The economics of implementing new clinical pathways across community and hospital-based care

Rua, Tiago

Awarding institution:
King's College London

The economics of implementing new clinical pathways across community and hospital-based care.

Tiago Duarte de Oliveira Rua

King's Health Economics

PhD in Health Service and Population Research at King's College London

April 2020

ABSTRACT

Imaging is the foundation of almost all clinical pathways and an increasingly influential tool for both the diagnosis and treatment of a wide spectrum of conditions. This thesis describes three individual studies, one randomised controlled trial (RCT) and two observational studies, performed in the context of an organisation-wide transformation initiative called TOHETI (**T**ransforming **O**utcomes and **H**ealth **E**conomics **T**hrough **I**maging) at Guy's and St Thomas' NHS Foundation Trust (GSTT). All three studies converged in the evaluation of the effectiveness and cost-effectiveness associated with the innovative use of advanced imaging in the context of clinical pathways in the National Health Service (NHS).

The economics associated with the utilisation of advanced imaging has changed over time due to the combined effect of growth in demand and technological developments that have led to advanced imaging becoming more accurate, accessible and less costly. However, there is still limited evidence around the cost and health economic implications associated with the use of advanced imaging. A systematic review conducted by the student assessed the challenges and methodological approaches used in the economic evaluation of diagnostic tests and constituted the foundation for the study design and statistical analyses employed in all three empirical studies.

The first study, a single-centre RCT, assessed the immediate use of Magnetic Resonance Imaging (MRI) in the management of suspected scaphoid fractures in the emergency department (ED) at GSTT. This study followed a published systematic review conducted by the student that highlighted the lack of both empirical economic evidence and appropriate economic modelling evidence on the immediate use of advanced imaging in the management of suspected scaphoid fractures. One-hundred and thirty six participants entered the study and were randomised to receive either the intervention with immediate MRI or follow routine care, which did not consider the use of MRI in the acute setting. This study was truly innovative as, to our knowledge, MRI is not considered in the context of acute care. The primary outcome was to estimate the cost implications from the healthcare payer perspective. Secondary outcomes included wider costs, cost-effectiveness and cost-utility (cost per quality-adjusted life years), diagnostic accuracy, clinical findings, time taken to reach a definitive diagnosis and patient satisfaction. Generalised linear models (GLM) were undertaken to estimate the main effect of group in all cost analyses. Based on intention-to-treat principles, the use of immediate MRI led to cost savings and, given the available data, there was a 96% to 100% probability of being cost-effective at conventional willingness-to-pay thresholds in the UK.

The second study, a single-centre pragmatic observational study, evaluated the utilisation of GP direct access to MRI compared to referral to neurology services for patients with chronic headache. Despite the benign nature of most headaches, headache management is associated with high healthcare utilisation, accounting to up to one third of neurologist appointments. The study's underlying hypothesis was that the early use of an advanced and accurate diagnostic tool (in this case MRI) would reassure

both patients and GPs that no serious underlying cause (particularly brain tumour) was present. This would in turn reduce the headache burden and NHS resource use associated with the patient's subsequent management. For this purpose, a total of two-hundred and forty nine patients were recruited for both groups (MRI and neurology groups) as per standard care. The primary outcome was to estimate the cost implications from the healthcare payer perspective. Secondary outcomes considered further cost and cost-effectiveness analyses, accessibility to care, time off-work and patient satisfaction. Cost analyses were conducted using GLMs and, given the study's non-randomised design, adjusted for potential imbalances at baseline. Based on intention-to-treat principles, direct referral to brain MRI from primary care led to cost savings, quicker access to care but lower patient satisfaction levels when compared with referral to neurology services.

The third study, a single-centre pragmatic observational study, evaluated the utilisation of Computed Tomography Colonography (CTC) compared to Optical Colonoscopy (OC) as the first line colonic investigation in the assessment of patients with low to intermediate risk of colorectal cancer (CRC). CRC is one of the leading causes of mortality and morbidity worldwide, with the UK presenting five-year survival rates significantly lower compared to other countries. Recent clinical guidelines aimed to increase early diagnosis of CRC by lowering the threshold for colonic investigations. However, this led to a substantial increase in colonic investigations, particularly OC, the diagnostic reference test for CRC, which remains technically difficult and resource intensive. This study evaluated the substitution of CTC as a first-line colonic investigation for patients deemed at low to intermediate risk of CRC. The underlying rationale was that CTC, a non-invasive and less costly colonic investigation, would be able to rule-out CRC or large polyps, thereby avoiding the need for invasive OC tests. Moreover, this would release much needed OC resources to test patients with known CRC or at a higher risk of CRC. The primary outcome was to estimate the cost implications from the healthcare payer perspective. Secondary outcomes considered cost-effectiveness and cost-utility, accessibility to care and patient satisfaction. Based on intention-to-treat principles, the use of CTC generated cost savings and presented a probability of 84%-91% of being cost-effective at conventional willingness-to-pay thresholds. The use of CTC also improved access to care, with no impact in patient satisfaction.

The role of observed data versus economic modelling is discussed taking into consideration published economic literature and its implications to interventions in the medical imaging field. The findings from *a priori* decision-analytical models were then compared to the empirical evidence retrieved from the three studies. Additionally, the student investigated whether the two methodological approaches would have led to different decisions from policy makers and ultimately affect the adoption of medical imaging technologies.

The last chapter completes the thesis with an overarching discussion of the main findings from model and real-world studies and their implications in the wider context of real-world NHS clinical practice. The implementation plans for the three different clinical pathways are detailed with the aim of bridging the gap between the clinical and economic evidence and the actual delivery of care across the NHS.

This thesis is dedicated to all NHS patients.
May we find a better way to improve your care.

ACKNOWLEDGMENTS

There is a list of people that I would like to thank for either their help with the work presented in this thesis and/or for their friendship and support throughout this journey.

First and foremost, I would like to express my deepest gratitude to my supervisors, Dr James Shearer, Professor Vicky Goh and Professor Paul McCrone. Many thanks for all your help, mentorship and encouragement. It has been a long journey and I could not have made it without you. James, many thanks for your thesis proof-reading.

I would also like to deeply thank the support of four people, Jo Turville, Professor Reza Razavi, Professor Anita Patel and Professor Janet Peacock. Jo, more than just my line manager, has always genuinely encouraged and supported my personal development. Many thanks to Professor Reza Razavi for his guidance and mentoring throughout this PhD process. It truly was an inspiration to work with Jo and Reza from the inception of the TOHETI programme. I would also like to thank Professor Anita Patel, my initial supervisor, who strongly recommended that I pursued this academic opportunity. The PhD title is still a result of our discussions four years ago! Many thanks to Professor Janet Peacock for your knowledgeable statistical support, from prompt replies to many emails, to your availability and kindness in our meetings. It has always been a real pleasure to work with all of you!

The TOHETI programme was a really challenging transformation initiative but I was very fortunate to encounter so many helpful members of staff at GSTT who did their outmost best to facilitate this research. There is a long list of clinicians (doctors, nurses, radiographers), managers and administrative staff whom I would like to acknowledge, particularly in the emergency and radiology departments. However, at the risk of forgetting so many, I have opted to name the few directly involved with the design and conduct of the three research studies: scaphoid study (Chief Investigator: Mr Sam Gidwani and Principal Investigator: Dr Sanjay Vijayanathan); headache study (Chief Investigator: Dr Shazia Afridi and Principal Investigator: Dr Asif Mazumder); and colon study (Chief Investigator: Dr Nyree Griffin and Principal Investigator: Ms Harriet Watson).

I have also been very fortunate to have different members of the research team who helped me with data collection, particularly Hara, Bharti, Yvonne, Rose, Sophia and all the Emergency Nurse Practitioners. Many thanks as I could not have supported the recruitment of so many participants across multiple studies by myself.

A remarkable mention should also go to all more than one thousand five hundred participants who took part in the different TOHETI research studies. I was inspired by their bravery and commitment to helping others in such a difficult time for them. I hope this work will help improve NHS care management for other patients like you.

I am also grateful to my friends on the outside. With some I have shared the whole journey of this PhD, many I had to leave behind after coming to the UK, but they all have a very special place in my mind. Thanks for all your support Casa, Che and Joaquim!

I am thankful to my family: my wonderful parents and brother. Thank you particularly for believing in me and encouraging me to pursue my goals. My uncle, who is no longer with us, remains a true inspiration for his kindness and genuine happiness.

And Liliana… how can I even begin to thank you? Without you, I would not have moved abroad and pursued this amazing opportunity. But, moreover, thank you for always supporting me and being the best person I have ever and will ever meet.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF EQUATIONS

LIST OF ABBREVIATIONS

ACE	Academic Centre for Evidence-Based Practice
CCG	Clinical Commissioning Group
CEAC	Cost-Effectiveness Acceptability Curve
CFIR	Consolidated Framework for Implementation Research
CHEERS	Consolidated Health Economic Evaluation Reporting Standard
CRC	Colorectal Cancer
CRF	Case Report Form
CT	Computed Tomography
CLIMP	Clinical Imaging and Medical Physics
CONSORT	Consolidated Standards of Reporting Trials
CRIS	Clinical Record Interactive System
CTC	Computed Tomography Colonography
DAP	Diagnostics Assessment Programme
DNA	Did Not Attend
ED	Emergency Department
ENP	Emergency Nurse Practitioners
EPR	Electronic Patient Record
EUnetHTA	European Network for Health Technology Assessment
FDA	Food and Drug Administration
FP	False Positive
FN	False Negative
GLM	Generalised Linear Model
GP	General Practitioner
GSTT	Guy's and St Thomas' NHS Foundation Trust
HADS	Hospital Anxiety and Depression Scale
HCHS	Hospital & Community Health Services
HES	Hospital Episode Statistics
HIT-6	Headache Impact Test 6
HRG	Healthcare Resource Groups
ICER	Incremental Cost-Effectiveness Ratio
ISPOR	International Society for Pharmacoeconomics and Outcomes Research
IT	Information Technology
ITT	Intention to Treat
KCH	King's College Hospital
MICE	Multiple Imputation using Chained Equations

MIDAS	Migraine Disability Assessment
MRI	Magnetic Resonance Imaging
MRSA	Methicillin-resistant Staphylococcus aureus
MTEP	Medical Technologies Evaluation Programme
NHS	National Health Service
NHSCR	National Health Service Central Register
NICE	National Institute for Health and Care Excellence
OC	Optical Colonoscopy
OECD	Organisation for Economic Co-operation and Development
PARIHS	Promoting Action on Research Implementation in Health Services
PDSA	Plan, Do, Study, Act
PET	Positron Emission Tomography
PICOS	Population, Intervention, Comparator, Outcome, Study
PIMS	Patient Information Management System
PPI	Patient and Public Involvement
PSS	Personal and Social Services
QALY	Quality Adjusted Life Year
RCT	Randomised Controlled Trial
RE-AIM	Reach, effectiveness, adoption, implementation, maintenance.
RWE	Real-World Evidence
SAP	Statistical Analysis Plan
SD	Standard Deviation
SMaRT	Scaphoid Magnetic Resonance Imaging in Trauma
TOHETI	Transforming Outcomes and Health Outcomes Through Imaging
TFCC	Triangular Fibrocartilage Complex
TP	True Positive
TN	True Negative
UK	United Kingdom
US	United States
VAS	Visual Analogue Scale

Chapter 1. Introduction

1.1. Context and rationale

Medical imaging encompasses diverse imaging modalities, from radiograph-based examinations to advanced imaging modalities such as Computed Tomography (CT) and Magnetic Resonance Imaging (MRI). Medical imaging is the foundation of many clinical pathways, being used in the diagnosis, monitoring or treatment of new or existing clinical conditions. The demand for medical imaging, particularly advanced imaging modalities, is growing across developed healthcare systems. At the same time, these healthcare systems are under increasing financial pressures. To date, there has been limited economic evidence around the use of medical imaging in real-world patients and clinical pathways.

This thesis aims to fill the gap between the economics and the medical imaging fields. It will examine whether the innovative use of advanced imaging may not only improve clinical outcomes and patient satisfaction but also, and despite the higher initial costs, could hold the potential to contribute to the National Health Service (NHS) financial sustainability agenda. This will be achieved using a detailed and consistent health economics methodology across different clinical conditions and imaging modalities. Furthermore, the use of real-world patients and clinical pathways, as opposed to decision analytical modelling only, will provide observed data on which to base potential changes in clinical practice.

This PhD was undertaken in the context of a wide transformation initiative - TOHETI (Transforming Outcomes and Health Outcomes Through Imaging) - at Guy's and St Thomas' NHS Foundation Trust (GSTT), a Tertiary Hospital in Central London. This programme was fully funded by the Guy's and St Thomas Charity (£13 million over a 4-year period). The overarching aim of this transformation programme was to evaluate the innovative use of medical imaging applied to real-world clinical pathways.

The TOHETI programme consisted of four phases (Figure 1). Phase 1 entailed the identification and selection of imaging-based transformational initiatives across the Trust. This process aimed to include any initiative that relied on the new or novel use of advanced imaging for the holistic transformation of the entire clinical pathway. Phase 2 included the proposal submission to secure funding. This process was based on *a priori* economic evaluation using decision analytical modelling. Phase 3 comprised the research studies which evaluated whether the intervention had the anticipated effects on costs and outcomes. This included preparation (i.e. study design, development of supporting documents, securing ethical and health research authority approval) and subsequent data collection and analysis and dissemination of findings. Lastly, phase 4 encompassed the implementation of the proposed initiatives as part of the normal clinical practice for patients at GSTT. This phase was dependent on

the findings from the research studies and the ability to overcome operational challenges associated with the proposed intervention across different clinical specialties and directorates.

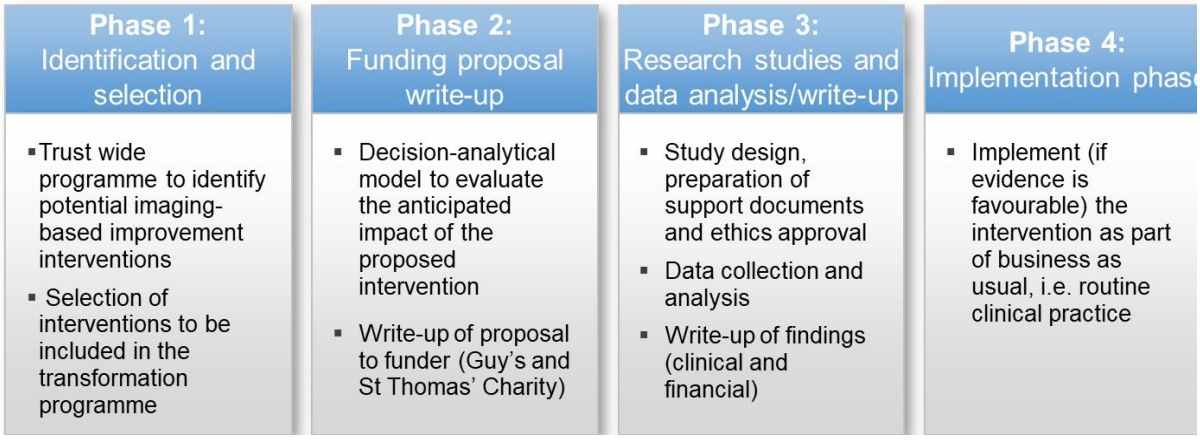


Figure 1. Four phases associated with the TOHETI programme.

As part of phase 1 of the TOHETI programme, and from over twenty potential initiatives, a total of six projects were selected (illustrated in Figure 2).

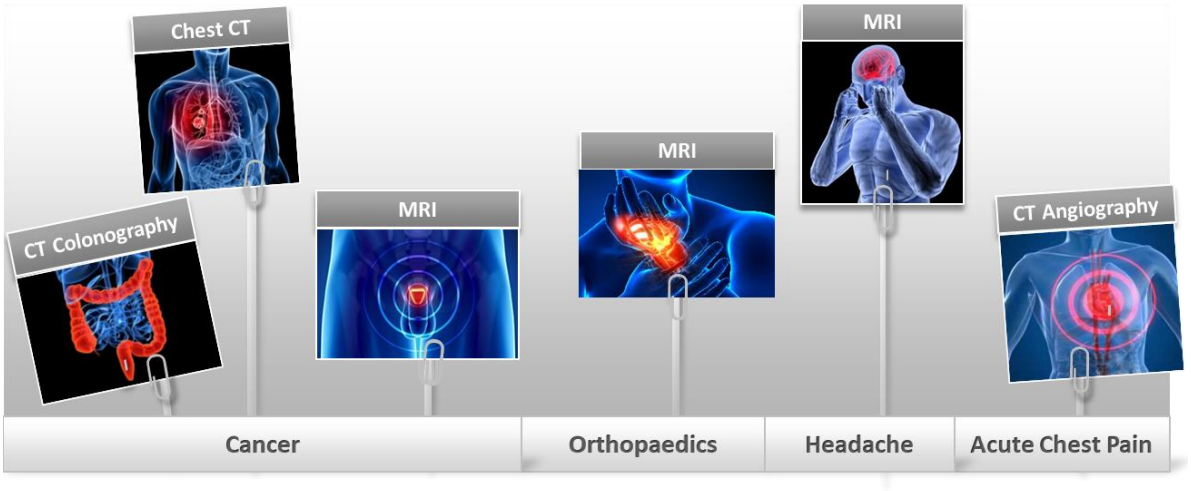


Figure 2. Six projects evaluated as part of the TOHETI programme.

Out of these six projects, **three are included in this thesis**:

1. The use of MRI in the acute management of suspected **scaphoid fractures** compared to the diagnostic strategy based on conventional radiograph only.
2. The direct referral from primary care to MRI for the management of **chronic headache** compared to the standard management of referral to a neurology clinic.
3. The use of CT colonography as the first line imaging investigation in the management of patients with low-risk **colorectal cancer** compared with optical colonoscopy.

The selection of initiatives was undertaken based on three criteria: (i) the alternative use of advanced imaging; (ii) high impact initiative and/or high volume clinical pathway; and (iii) the holistic transformation of diagnostic and treatment pathways. These criteria are explained in more detail below.

(i) All three projects involved the use of first-line advanced imaging in the diagnostic pathway. The suspected scaphoid fracture study investigated the use of MRI in the Emergency Department setting. To our best knowledge, this approach was unique worldwide. Likewise, the research involving patients with suspected colorectal cancer involved the use of non-invasive advanced imaging as a direct alternative to the invasive standard care optical colonoscopy in low-risk patients. The study with chronic headache patients investigated the use of advanced imaging as a direct alternative to a referral to a clinical specialist.

(ii) The suspected scaphoid fracture project involved the implementation of a high impact initiative, capable of affecting both clinical and efficiency outcomes given the superior diagnostic performance of wrist MRI compared to plain radiographs (Yin et al. 2010). The chronic headache and colorectal cancer pathways represented two high volume clinical pathways. In fact, chronic headache management, albeit being mainly managed within primary care, is the most common cause for referral accounting for up to 22% of GP referrals to neurologists (Thomas et al. 2010). According to NICE CG131, colorectal cancer is the third most common cancer in the UK (after breast and lung) and the second most common cause of cancer death in the UK (NICE 2011). Additionally, the UK 5-year survival rates for colorectal cancer are significantly lower than other countries (Coleman et al. 2011), with one of the possible contributing factors being the lack of timely diagnostic scans.

(iii) A holistic approach was present in all three research studies as the proposed interventions considered the transformation of the entire diagnostic and treatment pathways as a direct consequence of the advanced imaging findings. This meant that the innovative use of advanced imaging was the driver for change across the entire clinical pathway. As an example, a proportion of participants with no clinical findings in the wrist MRI will not require any follow-up at all. Hence, by using a more advanced and accurate imaging modality earlier in the clinical pathway, we will be able to change the participant's subsequent care. Similarly, a proportion of patients with chronic headache directly referred for a head MRI will not require a subsequent formal medical consultation with a neurologist.

1.2. Aims and questions addressed

This thesis applied methods of economic evaluation to assess three advanced imaging-based interventions in real-world patients across three different NHS clinical pathways. The overall aim of this thesis was to evaluate whether the innovative use of advanced imaging led to lower total costs from a healthcare payer perspective, thereby contributing to the NHS financial sustainability agenda. Second, the effectiveness and cost-effectiveness of the proposed interventions were also examined using

methodologies consistent with those recommended by the National Institute for Health and Care Excellence (NICE). These aims were then translated into individual research objectives organised by clinical study since the use of advanced imaging will lead to different findings in varied clinical contexts.

Three specific questions were addressed:

1. What are the economic and clinical benefits of using immediate wrist MRI in the acute management of suspected scaphoid fractures?

2. What are the economic and clinical benefits of using direct referral from primary care to head MRI in the management of patients with chronic headache?

3. What are the economic and clinical benefits of using CT colonography (virtual colonoscopy) in the management of patients with low risk suspicion of colorectal cancer?

1.3. Thesis structure

This thesis consists of seven chapters, organised around the three clinical studies completed as part of the PhD programme.

The present chapter (chapter 1) introduces the thesis and summarises the rationale around the selection of each intervention.

Chapter 2 provides background information around different diagnostic imaging modalities, with particular emphasis on advanced imaging. Considerations around the past, present and future utilisation rates, and its potential implications to the healthcare system, are summarised. The main challenges and methodological approaches used in the economic evaluation of diagnostic tests are also included in this chapter. Expanding on this, a systematic literature review was performed to assess the historical evolution and critically appraise different economic evaluation frameworks applied to diagnostic tests. Finally, the assessment of the published economic evidence of medical imaging technologies over a period of thirty years was performed using two existing systematic literature reviews (the most recent one dating from 2018).

Chapters 3, 4 and 5 are written as standalone chapters representing each individual research study, individual introduction, methods, results, discussion, conclusion and reference sections included within each chapter. Chapter 3 describes a randomised clinical trial assessing the innovative use of MRI in the acute management of suspected scaphoid fractures. Given the randomised design in the context of an acute pathway, particular attention was given to the challenges associated with this type of research. Chapter 4 investigates the use of MRI in the management of one of the most common clinical conditions, chronic headache, using non-randomised observational cohorts. Chapter 5 explores the use of CT as first line investigation in the assessment of patients with suspected colorectal cancer.

Chapter 6 compares the findings from *a priori* decision-analytical models to real-world prospective studies in the evaluation of specific medical imaging interventions applied to clinical pathways within the NHS. The role of observed data versus economic modelling is discussed taking into consideration published economic literature and the implications for interventions in the medical imaging field. Additionally, chapter 6 assesses whether the two methodological approaches would have led to different decisions from policy makers and, ultimately, affect the adoption of medical imaging technologies.

Chapter 7 completes the thesis with an overarching discussion of the main findings from the economic models and subsequent real-world studies and the implications in the wider context of the real-world NHS clinical practice. The implementation plans for the three different clinical pathways are detailed with the aim of bridging the gap between the clinical and economic evidence and the actual delivery of care across the NHS. Lastly, chapter 7 includes the considerations around the limitations and strengths of the work performed and recommendations for further research.

1.4. Student contribution

The current PhD was conducted in the context of a wide transformation initiative at a Tertiary Hospital in Central London - GSTT - and funded with £13 million by a local charity - Guy's and St Thomas' Charity. The idea for each clinical studies was chosen from over twenty advanced imaging initiatives proposed by clinicians from different clinical areas. The student, with special interest and academic background in medical imaging, was included in the process of triage and discussion of the proposed interventions. Subsequently, the student was involved in the write-up of the grant proposal, drafting the five clinical studies to be performed and the anticipated clinical and economic benefits based on decision-analytical models. As part of this thesis, three studies were selected whilst three other studies were not included (around the use of CT in acute chest pain, the feasibility of using CT in the diagnosis of lung cancer among asymptomatic participants and the use of MRI in the management of suspected prostate cancer).

Once the initiatives to be included as part of TOHETI were selected, the student developed the methodological approaches in coordination with the academic supervisors (Professor Paul McCrone, Professor Vicky Goh and Dr James Shearer) and statistical experts (Professor Janet Peacock and her team). All documents (e.g. study protocol, patient information sheet and other supporting documents) for all clinical studies were written by the student. Moreover, given the inclusion of real-world patients, all studies were centrally submitted for ethical approval (via the Integrated Research Application System). This entire process was conducted by the student in close coordination with different stakeholders (Chief Investigator, Principal Investigator and other NHS professionals - e.g. doctors, nurses, radiographers, managers).

Given the complexity of each study and the number of participants recruited simultaneously across different medical specialties (over 500 participants), participant data were mainly collected by research assistants and members of the clinical care team (e.g. emergency nurse practitioners, research nurses). The student was responsible for enrolling a small proportion of participants (~10%), and was responsible for providing training and continuously supporting all recruiters across the different studies. For one study, this meant, for instance, attending to regular handover meetings in the Emergency Department.

All collected data were manually entered into a web-based Case Report Form software (RedCap) by the student or research assistants. All database maintenance and data cleansing was carried out by the student. All statistical analyses of clinical and economic data described in this thesis were carried out by the student with support from the academic supervisors and statistical experts. The student wrote the first draft of the thesis which then was circulated among all academic supervisors who provided the student with invaluable comments and suggestions and proofread the final version of the thesis.

The student would like to highlight that the scientific approach associated with the TOHETI programme was well received by both managers and clinicians alike as a methodology to assess clinical pathways. This led to the inception a new Trust wide transformation initiative, called Care Redesign, which embodies the approach summarised in this thesis. At the time of writing of the thesis, over sixty multidisciplinary teams have looked at their operating systems and are aiming to them using some of the methods implemented in this PhD (e.g. cost analyses, economic modelling). Apart from the clinical impact of the three studies considered, this is one of the major legacies resulting from this PhD.

1.5. Publications and conferences

The findings from the three studies included in this PhD have resulted in the following peer-reviewed publications (listed in chronological order from the oldest to the most recent):

1. **Rua T**, Vijayanathan S, Parkin D, Goh V, McCrone P, Gidwani S. Rationale and design of the SMaRT trial: a randomised, prospective, parallel, non-blinded, one-centre controlled trial to evaluate the use of Magnetic Resonance Imaging in acute setting in patients presenting with suspected scaphoid fracture. *Journal of Clinical Trials*, 2017.

2. **Rua T**, Gidwani S, Parkin D, Goh V, McCrone P. The economic evidence for advanced imaging in the diagnosis of suspected scaphoid fractures: Systematic review of evidence. *Journal of Hand Surgery*, January 2018.

3. **Rua T**, Malhotra B, Vijayanathan S, Hunter L, Peacock J, Shearer J, Goh V, McCrone P and Gidwani S. Clinical and cost implications of utilising immediate Magnetic Resonance Imaging (MRI) in the management of patients with suspected scaphoid fracture and negative

radiographs: results from the SMaRT trial. *The Bone & Joint Journal, vol. 101-B, no. 8, pp. 984–994, Jul. 2019.*

4. **Rua T**, Gidwani S, Malhotra B, Vijayanathan S, Hunter H, Peacock J, Goh V, McCrone P, Shearer J (2020). Cost and cost-effectiveness implications of utilising immediate acute Magnetic Resonance Imaging (MRI) in the management of patients with suspected scaphoid fracture: results from a randomised clinical trial in England. *Value in Health* (in press).

5. **Rua T**, Mazumder A, Akande Y, Margariti C, Ochulor J, Turville J, Razavi R, Peacock J, McCrone P, Goh V, Shearer J, Afridi S. The management of chronic headache with referral from primary care to direct access to Magnetic Resonance Imaging (MRI) compared to Neurology services: an observational prospective study. *BMJ Open* (in press).

6. **Rua T,** Watson H, Malhotra B, Margariti C, Turville J, Razavi R, Peacock J, McCrone P, Goh V, Shearer J, Griffin N (2020). An observational study to compare the utilisation of Computed Tomography Colonography with Optical Colonoscopy as the first diagnostic imaging tool in patients with suspected colorectal cancer. *Clinical Radiology* (in press).

In addition, the PhD work has been presented at different national and international conferences. The student was responsible for submission of all abstracts, being the presenter in health management and health economics conferences and delegating to the senior clinical author the presentation role in most clinical conferences. The list of conferences and type of presentation (oral presentation or poster presentation) are presented in chronological order, with the presenter identified with an asterisk:

1. **Rua T***, Vijayanathan S, Shearer J, Goh V, McCrone P, Gidwani S. Transforming healthcare using medical imaging as the driver for change (oral communication). *28th Congress of the European Association of Hospital Managers*, September 2018 in Lisbon, Portugal. A 'best presentation' award was granted to the student (out of over 100 oral communications).

2. **Rua T***, Malhotra B, Vijayanathan S, Hunter L, Sharer J, Peacock J, Goh V, McCrone P, Gidwani S. Clinical and cost implications of utilising immediate MRI in the management of patients with suspected scaphoid fracture and negative radiographs (oral communication). *Annual Congress of the British Society for Surgery of the Hand*, April 2019 in Swansea, Wales.

3. **Rua T***, Malhotra B, Vijayanathan S, Hunter L, Sharer J, Peacock J, Shearer J, Goh V, McCrone P, Gidwani S. The use of advanced imaging as the driver for change across the suspected scaphoid fracture pathway: single-centre trial results (poster presentation). *International Forum for Quality and Safety in Healthcare*, March 2019 in Glasgow, Scotland.

4. **Rua T***, Gidwani S, Malhotra B, Vijayanathan S, Isaac A, Hunter L, Sharer J, Peacock J, Goh V, Shearer J, McCrone P. The use of advanced imaging as the driver for change across the suspected scaphoid fracture pathway: single-centre trial results (poster presentation).

International Conference of the International Society for Pharmacoeconomics and Outcomes Research, May 2019 in New Orleans, USA.

5. **Rua T**, Isaac A*, Malhotra B, Vijayanathan S, Hunter L, Sharer J, Peacock J, Goh V, Shearer J, McCrone P, Gidwani S. The scaphoid MR Imaging in Trauma (SMaRT) trial (oral communication). *Annual Congress of the European Society of Musculoskeletal Radiology*, June 2019 in Lisbon, Portugal.

6. **Rua T**, Afridi S, Akande Y, Margariti C, Turville J, Razavi R, Peacock J, Shearer J, Goh V, McCrone P, Mazumder A*. An observational study to evaluate the management of patients with chronic headache with referral from primary care to direct access to magnetic resonance imaging (MRI) compared to neurology services (oral communication). *42th Conference of the European Society of Neuroradiology*, September 2019 in Oslo, Norway.

7. **Rua T**, Vijayanathan S, Mak D*, Zavareh A, Isaac A, Malhotra B, Hunter L, Sharer J, Peacock J, Goh V, Shearer J, McCrone P, Gidwani S. Clinical and cost-effectiveness implications of utilizing immediate acute Magnetic Resonance Imaging (MRI) in the management of patients with suspected scaphoid fracture and negative initial radiographs: results from a randomized clinical trial (oral communication). A 'case of the day' award was granted to the presentation. *Annual Conference of the Radiological Society of North America*, December 2019 in Chicago, USA.

8. **Rua T**, Afridi S, Akande Y, Margariti C, Turville J, Razavi R, Peacock J, Shearer J, Goh V, McCrone P, Mazumder A*. An Observational Study to Evaluate the Management of Patients with Chronic Headache with Referral from Primary Care to Direct Access to Magnetic Resonance Imaging (MRI) Compared to Neurology Services (oral communication). A 'case of the day' award was granted to the presentation. *Annual Conference of the Radiological Society of North America*, December 2019 in Chicago, USA.

9. Turville J*, **Rua T**, Malhotra B, Cronin B, Akande Y, Razavi R. Transforming Healthcare and Outcomes Using Medical Imaging as The Driver for Change (TOHETI): Transformation Program in a Central London NHS Trust Services (oral communication). *Annual Conference of the Radiological Society of North America*, December 2019 in Chicago, USA.

10. **Rua T**, Watson H, Malhotra B, Cleary J*, Margariti C, Turville J, Razavi R, Peacock J, McCrone P, Goh V, Shearer J, Griffin N. An observational study to compare the utilisation of Computed Tomography Colonography with Optical Colonoscopy as the first diagnostic imaging tool in patients with suspected colorectal cancer (poster presentation). *European Congress of Radiology*, March 2020 in Vienna, Austria.

Chapter 2. Background

2.1. Evolution of medical imaging

Medical imaging is used to visualise the human body in order to diagnose, monitor or treat medical conditions. Medical imaging is used across a wide range of clinical conditions, from simple injuries to potential life-threatening conditions, as well as clinical pathways, from emergency to elective care.

Medical imaging includes a wide range of modalities, from basic technologies such as conventional radiograph (X-ray) and ultrasound, to more advanced modalities like Computed Tomography (CT), Magnetic Resonance Imaging (MRI) or Positron Emission tomography (PET). More recently, as illustrated in Figure 3, hybrid imaging modalities i.e. imaging that utilises more than one imaging modality simultaneously have been introduced into clinical practice (e.g. PET/CT or PET/MRI).

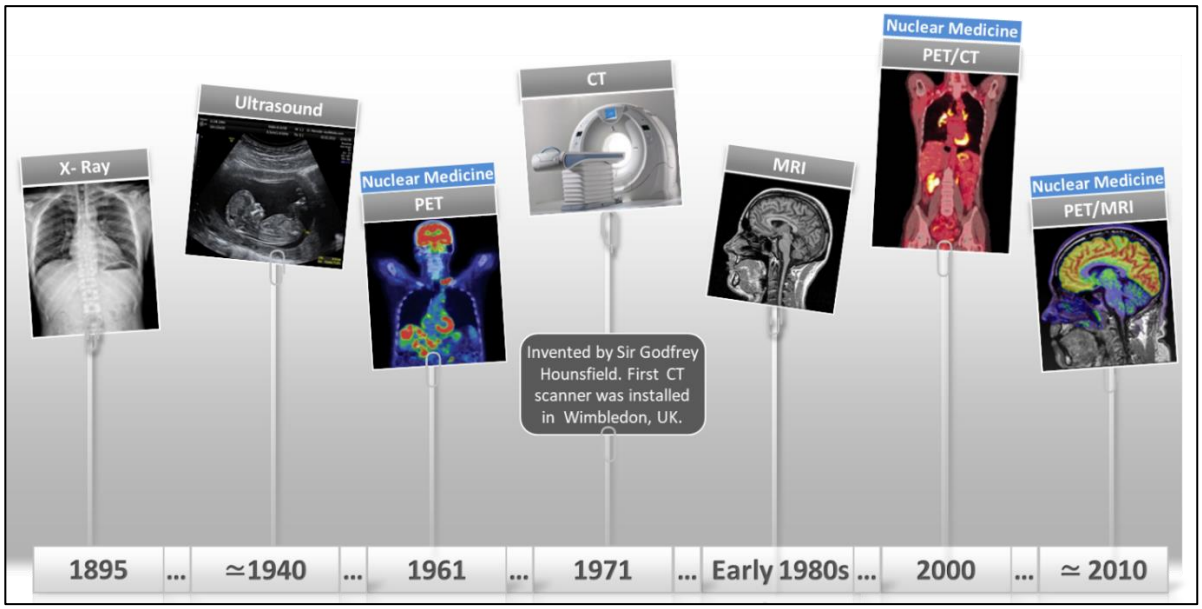


Figure 3. Key milestones in the development of the medical imaging field.

Medical imaging is an increasingly influential tool in the healthcare context. Furthermore, medical imaging is an evolving field, characterised by continuous improvement as well as disruptive innovations that hold the potential to revolutionise entire models of care. In fact, the increase in utilisation of medical imaging has historically surpassed the growth of the total healthcare expenditure, with the latest years however suggesting a slowdown in this trend in developed countries (Lang et al. 2013; Lee, Duszak, and Hughes 2013).

2.2. Past, present and future trends in medical imaging

2.2.1. Utilisation rates per inhabitant

As illustrated in Figure 4, the utilisation rates have increased over the last five years in England across all imaging modalities (Cake, Cavanagh, and Gordon 2015). This growth is particular relevant in the two advanced imaging modalities (CT and MRI), with annual growth rates of over 9%.

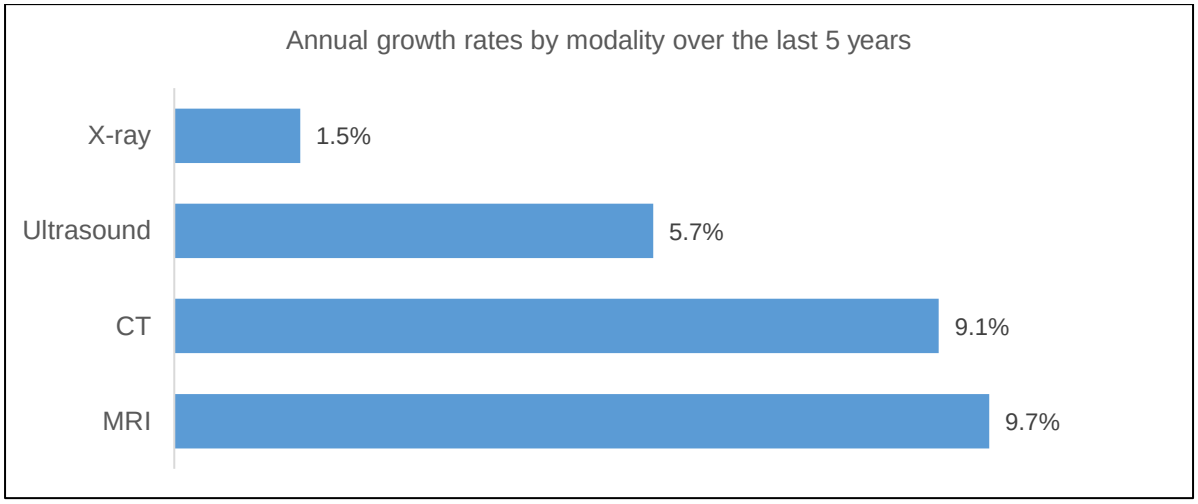


Figure 4. Annual growth rates by modality imaging over the last five years (Cake, Cavanagh, and Gordon 2015).

Additionally, Cake, Cavanagh, and Gordon (2015) estimated that the utilisation rates of advanced imaging modalities will increase even further, with projected annual growth rates over 10% between 2016 and 2020. This meant that in less than eight years, the utilisation rates for advanced imaging will have more than doubled. This growth of CT and MRI utilisation, illustrated in Figure 5, derives from demographic growth (estimated at 1.3% per year) but also from a change in clinical guidelines for the use of advanced imaging (for new or surveillance patients) and the lowering of clinical referral thresholds for a diagnostic test (e.g. direct access to imaging from primary care). In other words, it is the novel clinical use of advanced imaging modalities that is expected to drive the increase in utilisation rates of CT and MRI over and above the ageing population.

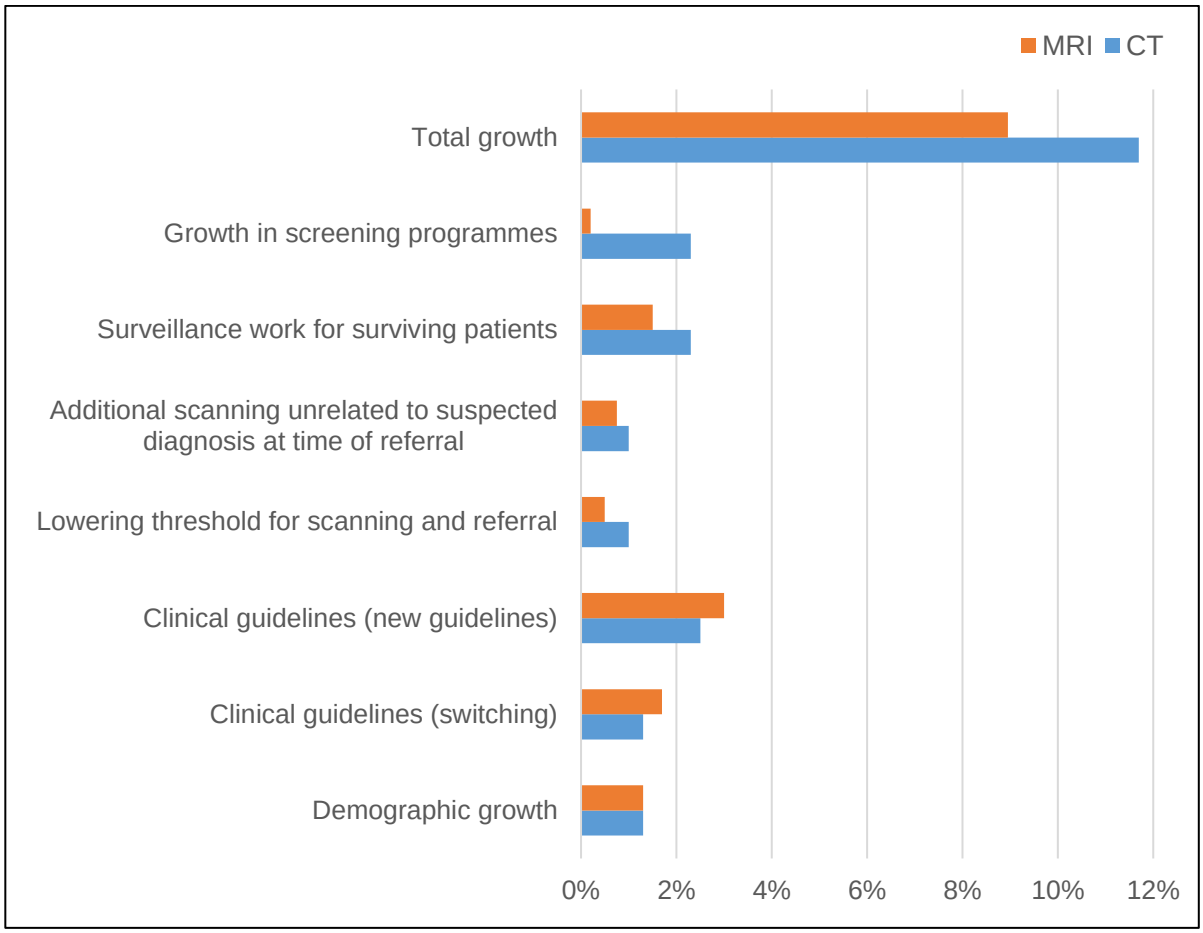
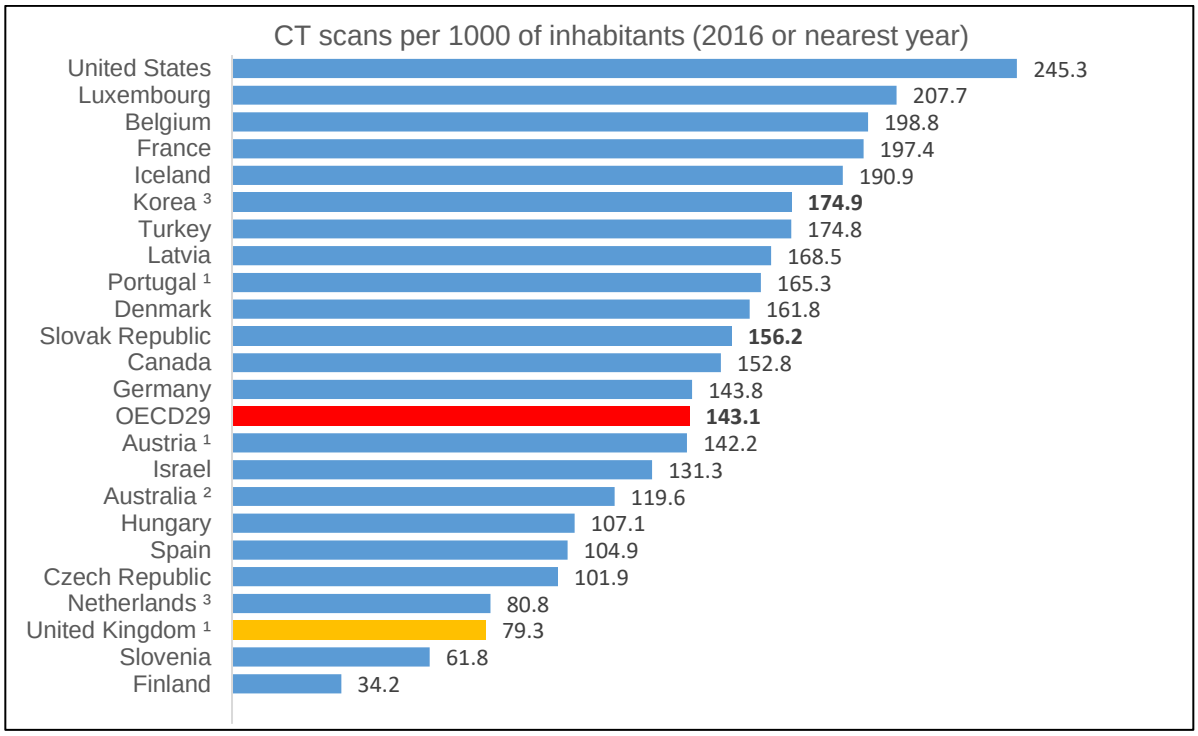


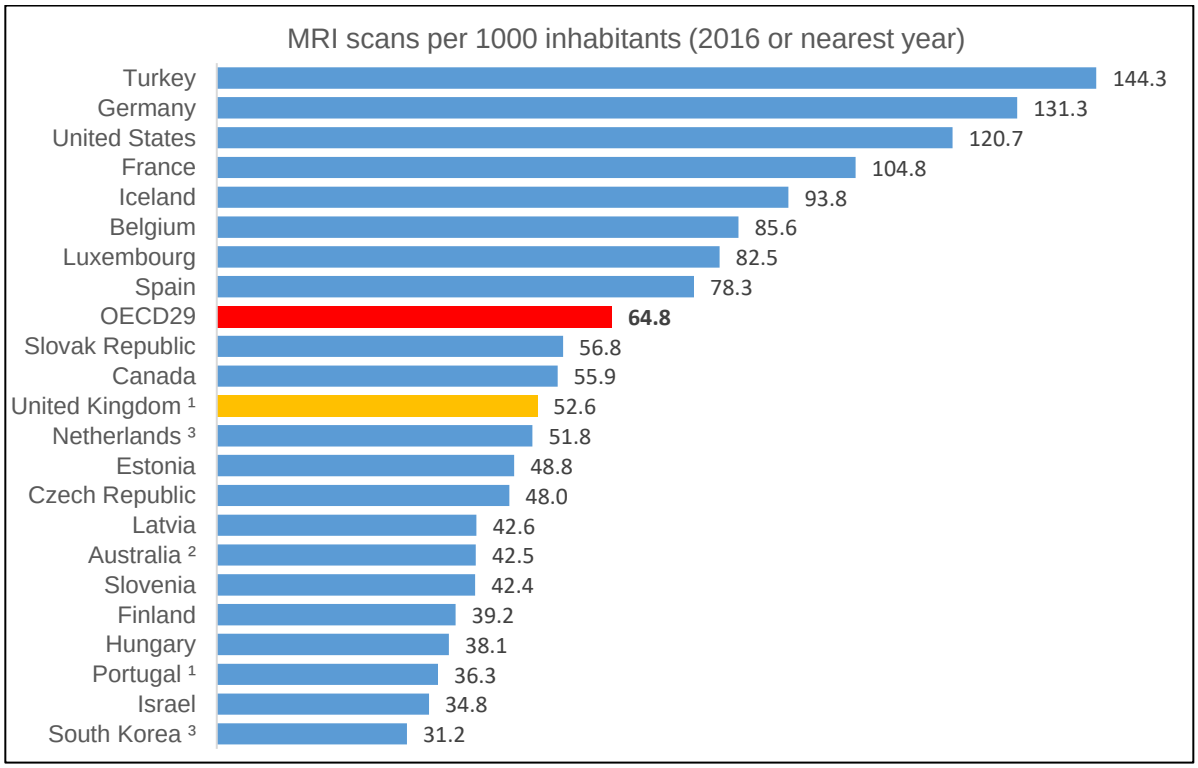
Figure 5. Expected annual growth rates for CT and MRI over the coming five years with the individual contribution of different factors (Cake, Cavanagh, and Gordon 2015).

These figures are relative to the baseline absolute utilisation rates for both CT and MRI in the United Kingdom (UK). In fact, the utilisation rates per inhabitant for both CT and MRI are significantly lower in the UK compared to other countries (see Figure 6 and Figure 7, respectively). The UK utilisation rates of CT and MRI in 2016 were, respectively, 45% and 19% lower than the average of the 29 countries considered from the Organisation for Economic Co-operation and Development (OECD) (see Figure 6 and Figure 7). In the case of CT, the UK is in the bottom three countries ranked by utilisation rates. This number however does not take into consideration exams performed outside hospital such as CT scans performed in private outpatient clinics but these are unlikely to significantly affect the overall UK position when compared to other developed countries.



Note: 1: CTs outside hospital excluded; 2: CTs on public patients excluded; 3: CTs privately-funded excluded.

Figure 6. Utilisation rate of CT per inhabitant in 2016 (or nearest year) (OECD 2018).

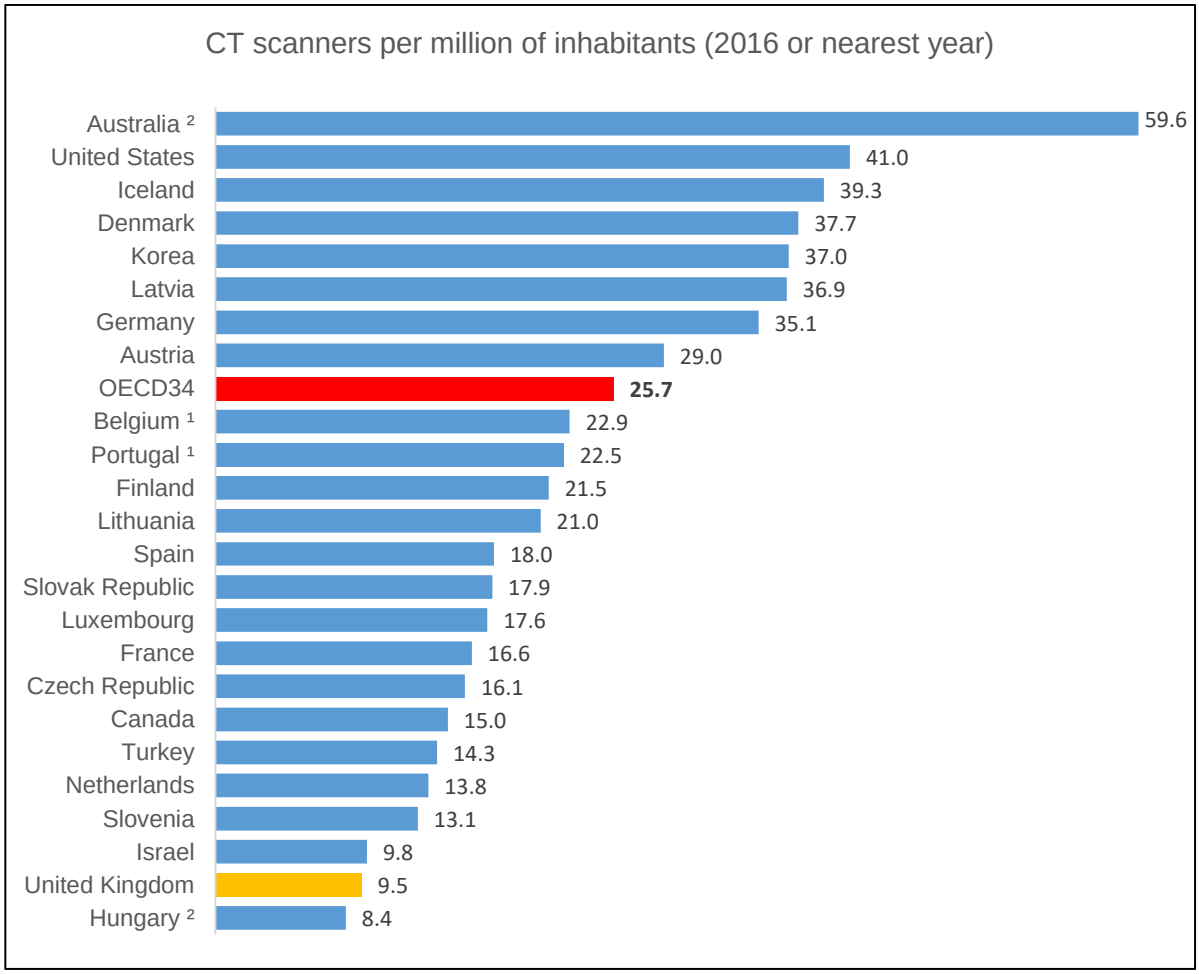


Note: 1: MRIs outside hospital excluded; 2: MRIs on public patients excluded; 3: MRIs privately-funded excluded.

Figure 7. Utilisation rate of MRI per inhabitant in 2016 (or nearest year) (OECD 2018).

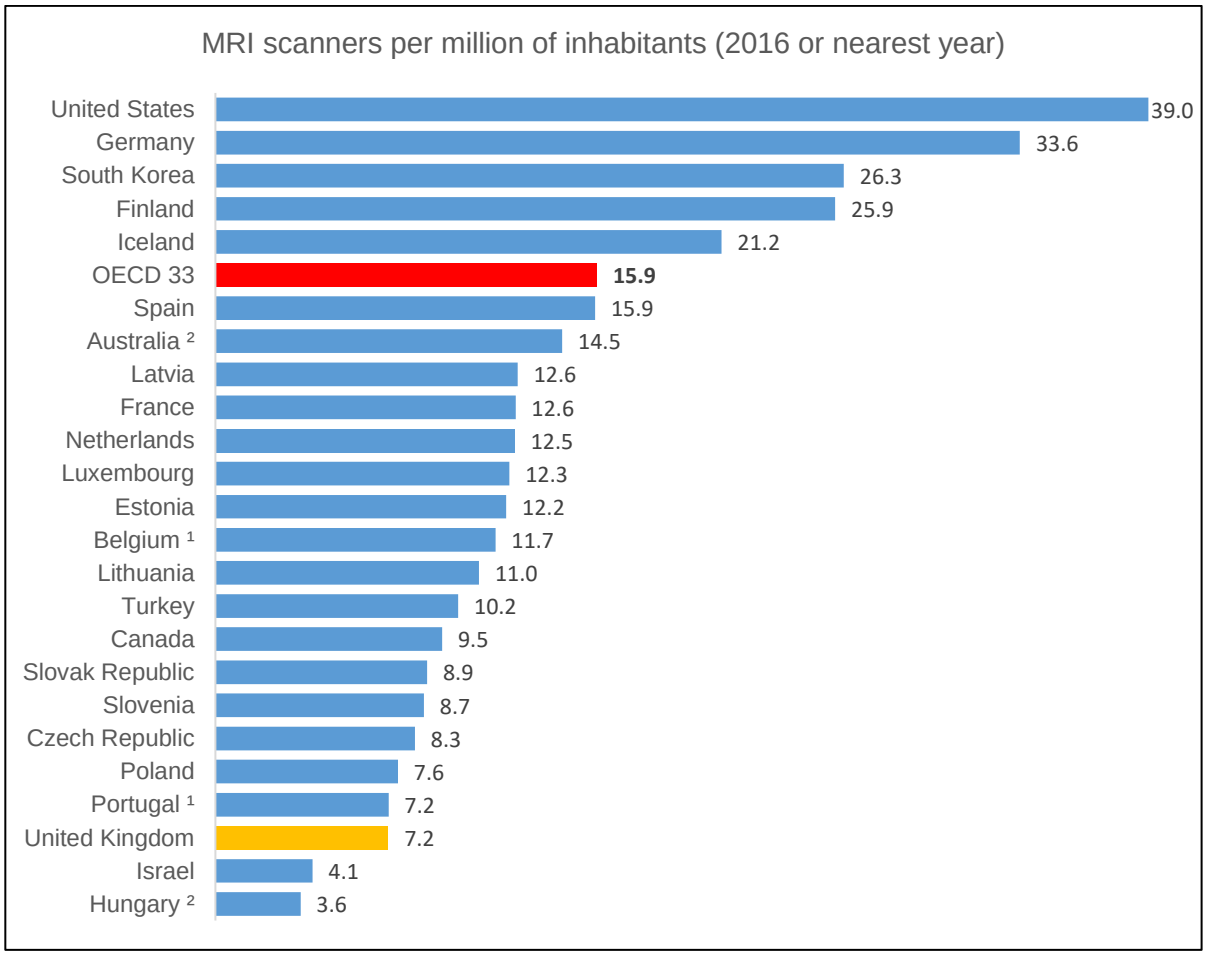
2.2.2. Number of scanners per inhabitant

The UK also has a lower number of both CT and MRI scanners per million inhabitants compared to the majority of OCDE countries. The UK had in 2016 fewer CT and MRI scanners, with a proportion 63% and 55% lower than the average of the 34 countries considered from the OECD (see Figure 8 and Figure 9), respectively. For example, the UK had a total of 9.5 CT scanners per million of inhabitants, much lower than the United States (US) (41.0), Germany (35.1) or France (16.6). Similarly, with regards to MRI, the UK had 7.2 scanners per million of inhabitants, again lower than the US (39.0), Germany (33.6) or France (12.6).



Note: 1- Equipment outside hospital excluded; 2- Only equipment eligible for public reimbursement.

Figure 8. Number of CT scanners per inhabitant in 2016 (or nearest year) (OECD 2018).



Note: 1- Equipment outside hospital excluded; 2- Only equipment eligible for public reimbursement.

Figure 9. Number of MRI scanners per inhabitant in 2016 (or nearest year) (OECD 2018).

2.2.3. Utilisation rate per scanner

A very important metric that is commonly ignored relates to the number of diagnostic tests per scanner. This metric is relevant as estimates the actual overall occupancy rates per scanner and hence is a proxy to the existing system capacity to accommodate any increase in demand. Although these data should not be seen in isolation, as it depends on different factors such as the type of healthcare system, other healthcare resources (e.g. radiologists) or even the countries' geography, the number of scans per scanner provides a solid indication of the current situation of advanced imaging in the NHS.

In the case of CT, the UK had a total of 8,383 CT scans per scanner, 20% higher than the average (6,999), 38% higher than the US (6,063) or 104% higher than Germany (4,098) but 28% lower than France (11,653) (Figure 10). A similar situation occurs with MRI, with the UK having 7,275 MRI scans per scanner, 32% higher than the average, 120% higher than the US (3,287) or 86% higher than Germany (3,904) but, again, 6% lower than France (7,740) (Figure 11).

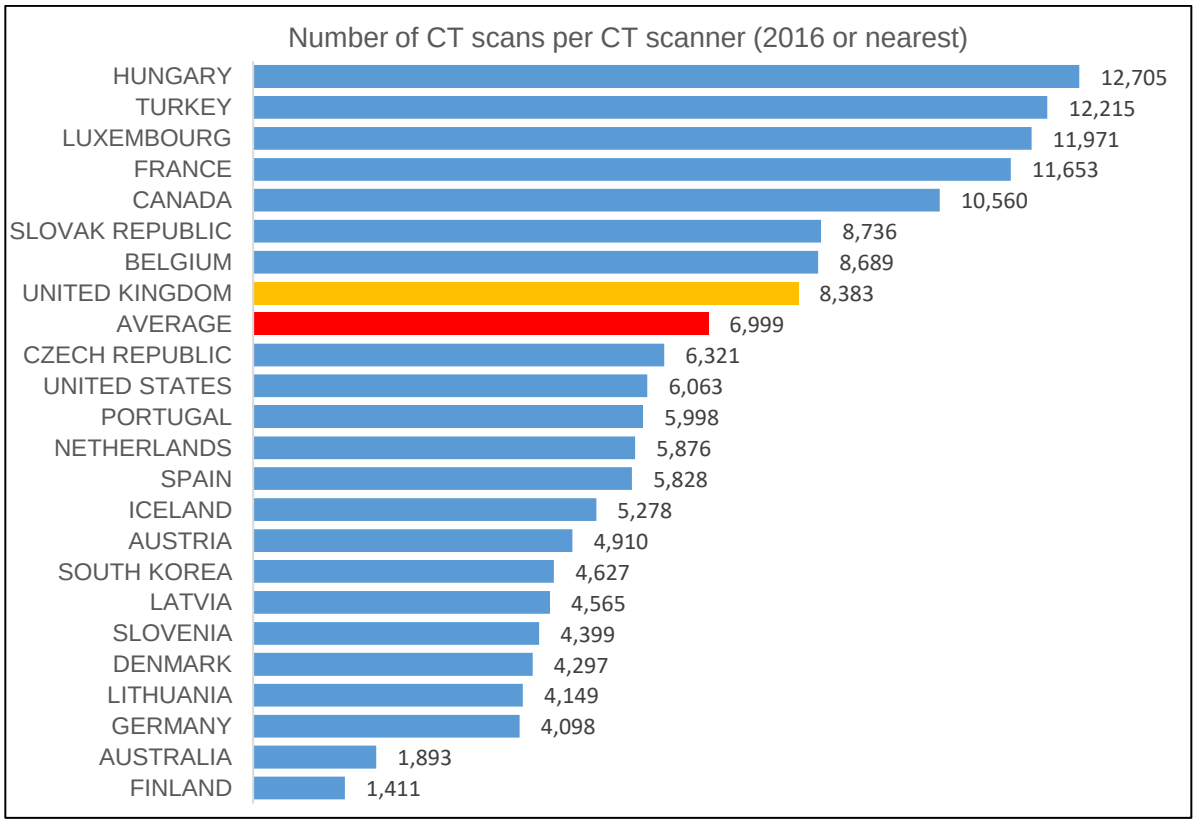


Figure 10. Number of CT scans per CT scanner in 2016 (or nearest year) (OECD 2018).

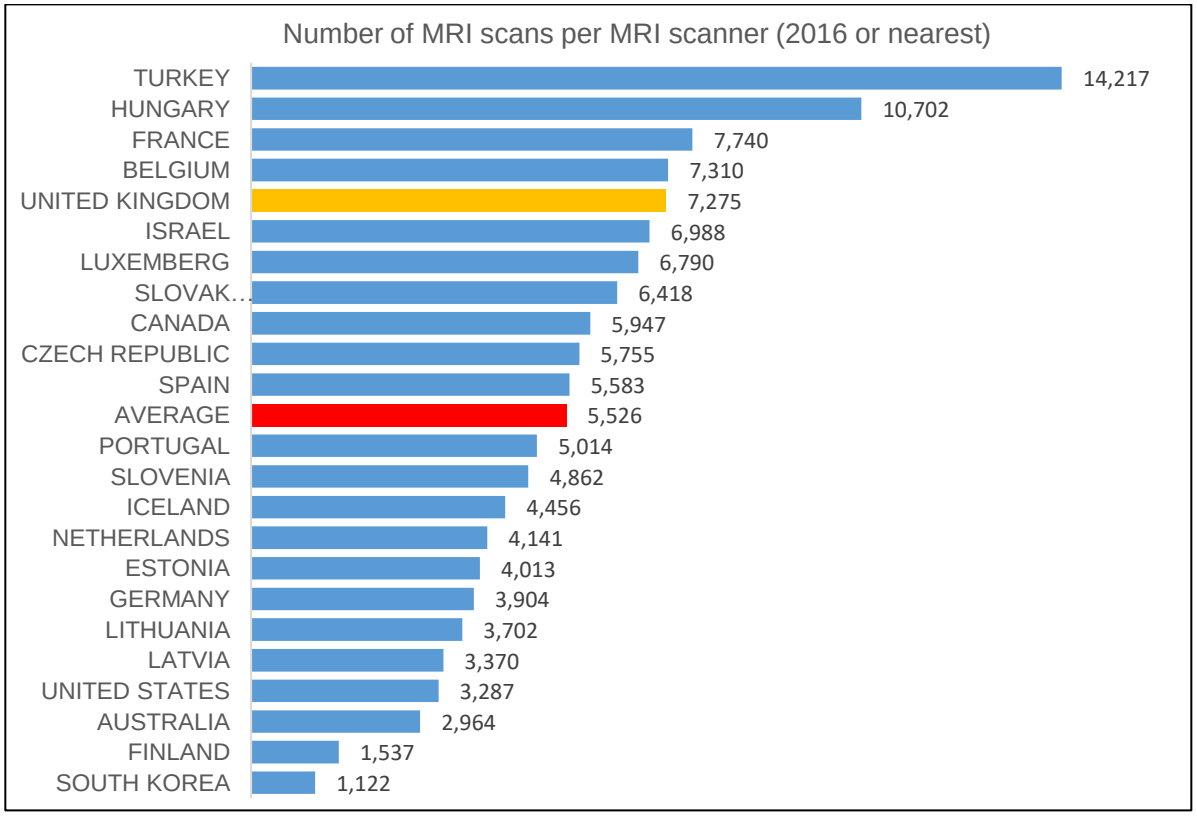


Figure 11. Number of MRI scans per MRI scanner in 2016 (or nearest year) (OECD 2018).

2.2.4. Advanced imaging in the UK: supply and demand

Figure 12 illustrates the background on the current and expected use of advanced imaging in the context of the UK healthcare market.

First, from the perspective of demand, the UK utilisation of advanced imaging (e.g. MRI, CT) is lower when compared with other countries, such as France, Germany or the US. Second, given the increasing use of advanced imaging in all healthcare systems and the UK utilisation gap when compared to these countries, the very high annual growth rates of advanced imaging in the UK are expected to continue in the near future. Hence, from a demand point of view, these two factors converge towards the expansion of advanced imaging in the NHS.

Third, from the supply perspective, the UK has a lower number of advanced imaging scanners (both CT and MRI scanners) when compared to other countries such as Germany, France or the US. Fourth, not only does the UK have a lower number of scanners per inhabitant, its occupancy rate is higher when compared to most countries. The latter means that the existing scanners are running to a higher occupancy rate and therefore have a limited ability to respond to an increase in workload. Thus, from a supply point of view, as the existing NHS capacity is not enough to accommodate the increasing demand needs, capital investment in the NHS will be needed over the coming years. This situation places an extra emphasis on the need for economic evaluation concerning the current utilisation of advanced imaging within the NHS.

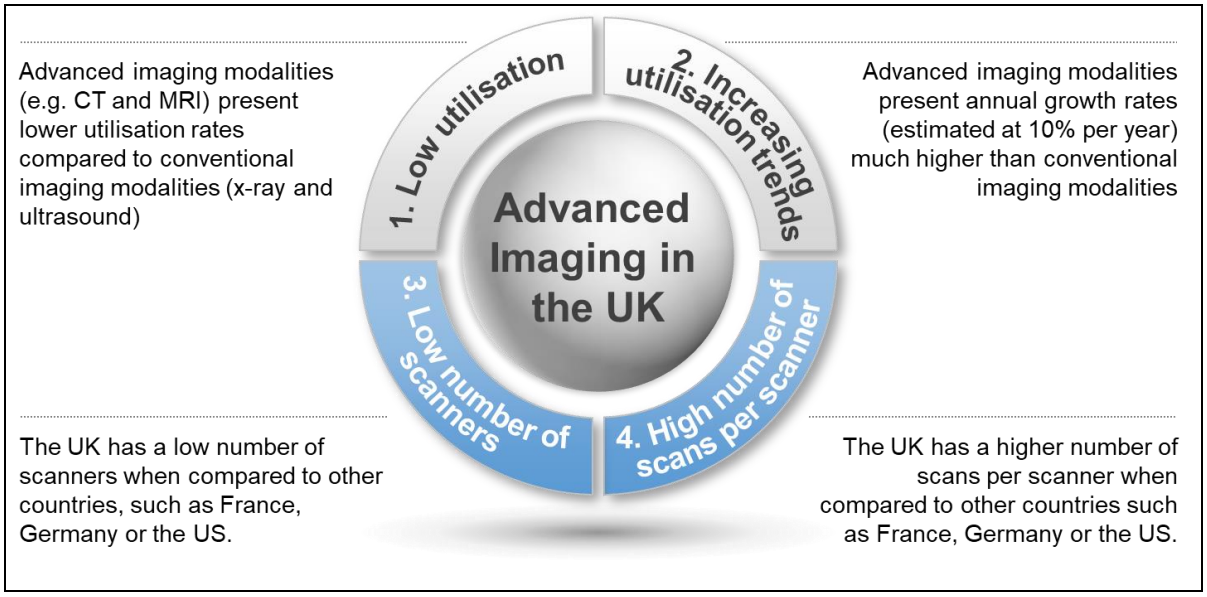


Figure 12. Demand and supply analysis associated with the use of advanced imaging the UK.

2.3. Economic evaluation of medical imaging

2.3.1. Rationale

Developments in the medical imaging field have simultaneously led to: (i) an increase in the accuracy of CT and MRI in the diagnosis of different clinical conditions; and (ii) a decrease in the acquisition and processing time associated with CT and MRI scans. The latter means that newer scanners enable more accurate and faster diagnostic scans, leading to an increased throughput per scanner (i.e. higher number of exams per scanner). Given that the acquisition costs of new CT and MRI scanners have remained relatively stable, the faster acquisition time per imaging test has led to a decrease in the operational unit cost per CT or MRI scan (European Society of Radiology, 2014). This trend is expected to continue over time. Hence, as advanced imaging becomes more accurate, accessible and cheaper, its use in the NHS needs to be reassessed. Although the purchase and implementation of new diagnostic equipment is associated with quicker access and improved diagnostic accuracy, there is limited evidence that this leads to improved health outcomes (Baker, Atlas, and Afendulis 2008). In conclusion, due to the increase in both the demand and the supply of medical imaging, there is an imperative need to establish economic criteria regarding its use in the context of the NHS (illustrated in Figure 13).

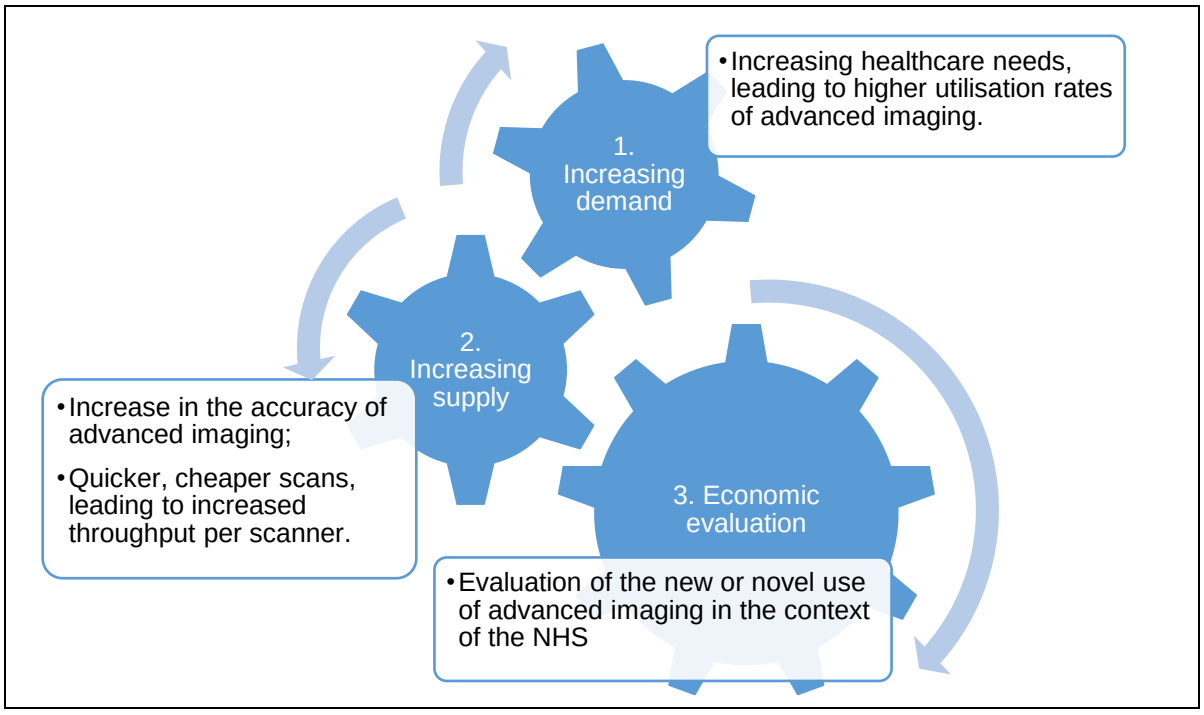


Figure 13. Implications of increasing demand and supply in the context of economic evaluation of advanced imaging.

2.3.2. Key challenges

It is recognised that there is more economic evidence around the evaluation of therapeutic (e.g. pharmaceutical drugs) than for diagnostic interventions (Drummond, Griffin, and Tarricone 2009). This is mainly due to a fundamental difference between assessing a therapeutic intervention (e.g. pharmaceutical drug or cardiac stent) and a diagnostic test (e.g. MRI). This difference is a cause and effect relationship (Hollingworth 2005). Whilst, for instance, the link between a pharmaceutical drug (the cause) and its impact at different outcomes (the effect) can be established in the context of a properly designed trial, the latter is harder with diagnostic tests (Drummond, Griffin, and Tarricone 2009). In fact, medical imaging, as other diagnostic tests, is used to generate diagnostic information of a potential or confirmed clinical condition. This diagnostic information may in turn impact the therapeutic decision, which can then have different outcomes (measure of effect). Figure 14 and the paragraphs below illustrate the five layers of uncertainty associated with the evaluation of value of diagnostic tests:

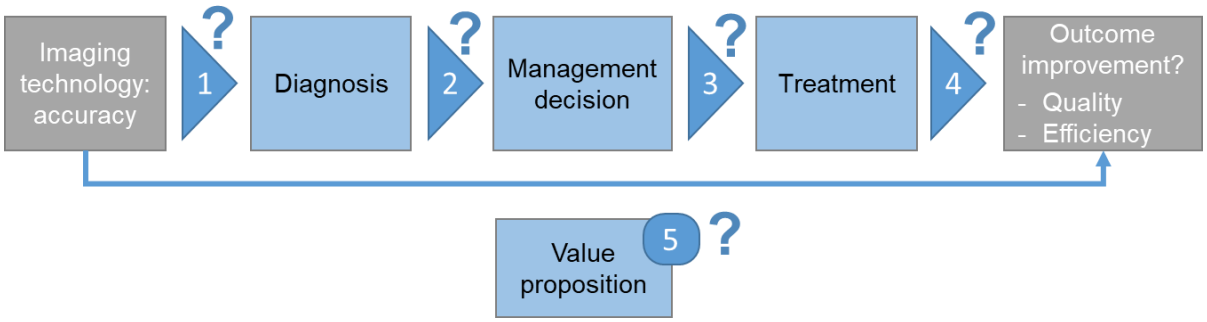


Figure 14. Uncertainty in the assessment of value in the use of a diagnostic tool.

- First, the rationale behind a newer and better diagnostic test is that it will improve the probability of achieving the right diagnosis. However, the accuracy of diagnostic tests is not perfect, given the presence of false positives (ruling in a disease in patients without the disease) and false negative findings (ruling out a disease in patients with the disease). Figure 15 illustrates a hypothetical scenario where the distribution of false positive (FP), true positive (TP), true negative (TN) and false negative (FN) findings is dependent on: (i) the disease prevalence or incidence (hypothesised at 20% in Figure 15); and (ii) the accuracy (hypothesised at 80% sensitivity and 90% specificity in Figure 15) of the diagnostic test for that condition. In this hypothetical clinical scenario, the diagnostic test led to the right diagnosis in 88% of the cases (16% TP + 72% TN) and the wrong diagnosis in 12% of the cases (8% FP + 4% FN). This introduces the first level of uncertainty, the test's ability to correctly identifying the disease in patients with the condition (sensitivity) and exclude the disease in patients without the disease (specificity).

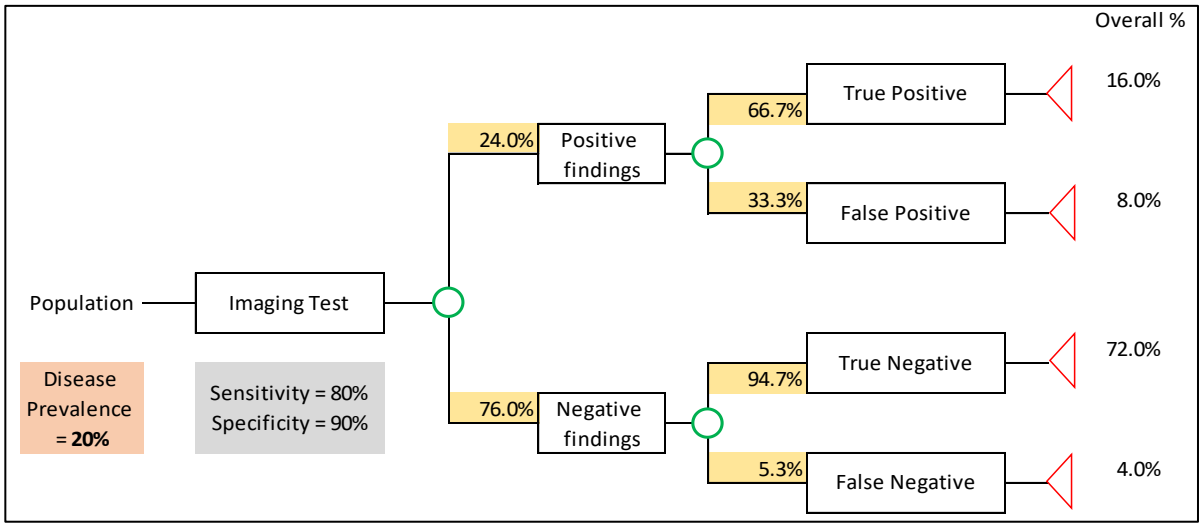


Figure 15. Decision tree associated with the use of an imaging test to a hypothetical cohort of patients.

- Second, even if the correct diagnosis is achieved, it will only be relevant if it affects the clinician's diagnostic decision. This can either be to change or to reassure the clinician's diagnostic thinking. The diagnostic test can also impact on patients' confidence in the diagnosis, affecting their perceptions and behaviours.
- The third layer of uncertainty relates to the probability of the right management decision or treatment (if any) being chosen. Indeed, the use of medical imaging might not lead to a change in the treatment options or, even if the treatment choice is impacted, there is uncertainty about whether the right treatment is chosen. Again, the diagnostic test can impact both the clinicians' and the patients' behaviours, e.g. reassuring the clinician that the right treatment is chosen or reassuring the patient and thereby improving their adherence to the treatment.
- The fourth layer of uncertainty is associated with the clinical effectiveness of the treatment itself. The same treatment applied to different patients with the same condition might result in different outcomes (clinical or otherwise).
- Fifth, there are several value propositions associated with diagnostic tests. A diagnostic test might be used as an add-on or replacement test in the context of specific clinical pathways or might be used for operational purposes (e.g. speed up the diagnosis), reduce costs or improve care. Furthermore, the same diagnostic technology applied in different clinical settings might have different value propositions and multiple applications (Drummond, Griffin, and Tarricone 2009).

Table 1 further details the main features associated with medical imaging tests that lead to the above mentioned five layers of uncertainty and pose challenges in the evaluation of imaging technologies.

Table 1. Key features responsible for the layers of uncertainty in the evaluation of medical imaging technology.

Uncertainty layer	Feature	Description
1 - Imaging technology accuracy / Diagnosis	1.1. Accuracy levels	The test's sensitivity defines its ability to identify the clinical condition in patients with that condition whilst its specificity refers to the test's ability to identify patients without the clinical condition. False positive and false negative results occur from the use of imaging tests which are not 100% accurate.
	1.2. Machine variability	The same imaging technology (e.g. CT) can be implemented in the same clinical condition using CT equipment from different companies. These machines are intrinsically different and this has the potential to affect the outcomes of the intervention.
	1.3. Reader variability	The same imaging scan can be reported differently by two readers. Given its subjectivity compared to other technologies (e.g. pharmaceutical drugs), this introduces uncertainty as the result from the reader's report (and not the images themselves) are the actionable information on which the patients' diagnosis and treatment might be changed.
	1.4. Reference test	The introduction of new imaging technology is usually compared against other imaging tests. The comparator tests are normally assumed to be the reference, with perfect accuracy (100% sensitivity and specificity) as part of decision-analytical models. This assumption is associated with uncertainty as no imaging test is actually 100% accurate.
	1.5. Incidental findings	The use of medical imaging tests might lead to the incidental identification of clinical condition(s) unrelated to the original request for the imaging test. Some of these incidental findings are clinically relevant and hence require follow-up and have the potential to affect the evaluation of the imaging test. However, the inclusion of incidental findings in decision-analytical models is typically ignored.

2 - Diagnosis / Management decision	2.1. Diagnostic decision	The diagnostic medical algorithm is usually based on clinical, laboratory and radiological findings. Hence, the imaging test's diagnosis is a central element that might affect the clinician's diagnostic thinking by either corroborating or rejecting the anticipated diagnosis.
	2.2. Clinician's reassurance	Even if the imaging test corroborates the known diagnosis, there is added value to the diagnostic test. In real-world clinical practice, clinicians request imaging tests for reassurance that the right diagnosis has been achieved. However, the added psychological value of the test is difficult to measure and evaluate.
	2.3. Patient's reassurance	Similarly to clinicians, patients' reassurance that the right diagnosis has been achieved might differ based on the imaging test. In fact, the use of imaging tests might impact on the patients' perception that the right/wrong diagnosis has been made and hence affect their behaviours.
3 – Management decision / Treatment	3.1. Management decision	Based on a given diagnosis, clinicians subsequently decide on the management/treatment strategy – if any at all (with do nothing as a possibility). Hence, the imaging test has contributed to a given diagnosis that, in turn, led to a specific treatment. However, the link between these events is often associated with uncertainty in real-world clinical practice.
	3.2. Clinician's reassurance	The use of imaging tests not only reassures the clinician that the right diagnosis was made but also that the management decision or treatment is appropriate.
	3.3. Patient's reassurance	The imaging test might affect the patients' perception that a given diagnosis and subsequent treatment is right or wrong, thus affecting their behaviours. One of these behaviours is adherence to treatment. If the patient is reassured that a given diagnosis/treatment is right, he/she is more likely to adhere to the treatment proposed by the clinician.
	Different dimensions of analysis	A medical imaging test can be used to optimise operational efficiencies (e.g. waiting times), streamline patient management (e.g. speed up diagnosis) and overall efficiency gains (e.g. reduce costs) and/or improve clinical outcomes (e.g. increased survival rates).

4 – Treatment / Outcome improvement	Surrogate references	Given the lack of long-term evidence, surrogate or intermediate references to evaluate clinical endpoints are commonly considered in economic evaluations. However, the statistical relationship and epidemiological causality between the surrogates and the clinical endpoints and the lack of standard in the use of surrogates adds to the uncertainty of evaluating diagnostic tests. As an example, diabetes is a known risk factor for cardiovascular events but there is little evidence that a tight control of diabetes (measured using the surrogate haemoglobin 1Ac) will reduce cardiovascular events (Weintraub, Lüscher, and Pocock 2015). Hence, the use of surrogate references as an alternative to empirical data from clinical studies, particularly randomised trials, can be problematic and misleading (Ciani et al. 2017).
5 - Value Proposition	5.1. Value proposition	The introduction of imaging tests in clinical pathways is associated with a value proposition such as: increased accuracy; faster and cheaper; ability to replace a more expensive test; early diagnosis.
	5.2. Functional utilisation	Imaging tests are used widely for: risk assessment, screening of asymptomatic patients, diagnostic test based on symptoms, monitoring the evolution of a known clinical condition (e.g. cancer surveillance), prognosis (e.g. risk stratification, measure response to cancer treatment) and support to treatment purposes (e.g. image-guided intervention).
	5.3. Role across clinical pathways	Imaging tests can be used as a triage, add-on or replacement tool in different clinical pathways. Furthermore, the same imaging test can be used differently in the context of varied clinical pathways.
	5.4. Multiple uses of the same test in different clinical pathways	The multiple uses of the same imaging test in different clinical contexts adds uncertainty because an imaging test might be cost-effective for a given clinical use and not cost-effective for another. Hence, in order to deal with this situation, the evidence needs to be somehow weighted to support/reject the introduction of such imaging technology.
	5.5. Real-world evidence vs decision analytical models	There is a lack of long-term real-world evidence of imaging tests from randomised trials. Decision-analytical models are typically used to circumvent this limitation but the lack of observed data adds uncertainty in the use of decision-analytical models.

Given the different levels of uncertainty associated with the evaluation of diagnostic tests, the assessment of such technologies is inherently complex when compared to other technologies (e.g. pharmaceutical drugs) (Drummond, Griffin, and Tarricone 2009). The next subsection summarises the evidence surrounding the use of evaluation frameworks specific to the evaluation of diagnostic tests (e.g. medical imaging) and appraises how these layers of uncertainty are to be addressed.

2.3.3. Evaluation frameworks to assess diagnostic tests

Whilst the framework for the evaluation of pharmaceutical drugs is well established as a four to five phase hierarchical model, the same does not happen in the evaluation of diagnostic tests (Gatsonis, 2012). Indeed, partly due to the intrinsic challenges associated with the evaluation of diagnostic tests, different conceptual models have been developed over time to support economic evaluations of these interventions.

A systematic review research paper (search run in January 2009) evaluated and described the different frameworks for the evaluation of medical tests (Lijmer, Leeflang, and Bossuyt, 2009). The study methods reported the search of key databases (Medline, Web of Science and Embase) with variations of the word “diagnostic” in the document and variations of the concept “phased model” in the title or abstract. This search criteria aimed to maximise the number of hits. A total of 19 models were identified, with the first one being published in 1978 and the most recent one in 2007. Models subsequent to the systematic literature review performed by Lijmer, Leeflang, and Bossuyt (2009) were included based on a second systematic literature review performed by the student using the same search terms and databases. This review further identified an additional four evaluation frameworks. Out of the 23 evaluation frameworks identified in the two systematic literature reviews, most were iterative improvements or amendments of previous frameworks. Although summarily described for completeness and better understanding of the continuous improvement historical approach, this section mainly focuses on the following four evaluation frameworks:

- Fryback and Thornbury (1991);
- Houn et al. (2000) from the Food and Drug Administration (FDA);
- Gazelle et al. (2011);
- National Institute for Health and Clinical Excellence (2011).

The selection of these specific four frameworks was based on novelty and impact both in terms of research and applicability to the real-world evaluation of diagnostic tests in the context of the NHS.

2.3.3.1. Fryback and Thornbury (1991): 6-tier evaluation framework

The first cohesive evaluation framework specific to diagnostic tests was presented by Fryback and Thornbury (1991) following on evidence from earlier manuscripts, with particular emphasis on Loop and Lusted (1978) and the equivalence between diagnostic studies and standard classification of

clinical trials proposed by Freedman (1987). The authors proposed a six-tier hierarchical framework of efficacy exclusive to diagnostic tests (Figure 16) (Fryback and Thornbury 1991). The first level of efficacy, described as technical efficacy, relates to the physical parameters assessing the technical quality of the image (e.g. number of lines per image or number of artefacts). The second tier of efficacy addresses the technology's diagnostic accuracy, typically described as the receiver operating curve or specificity/sensitivity. The model inherently considers that the uncertainty regarding the diagnosis is associated not only with the quality of the image (tier one) but also with the interpretation of such images by the reporter, i.e. the radiologist. The third tier, the diagnostic thinking efficacy, addresses the change in the diagnostic differential of the clinician due to information provided by the diagnostic test. In a given context, clinical outcomes are not affected by a diagnostic test unless that test leads to a change in the referrer's opinion. This level of efficacy is difficult to estimate without empirical research as diagnostic tests may not only change the referrer's decision but also strengthen or reassure him/her that a given diagnosis is correct. The fourth tier recognised the ability of the diagnostic test to actually impact the patient's clinical management (e.g. the percentage of times that the diagnostic test changed the subsequent clinical practice). Hence, the model considers that some tests might change the referrer's diagnosis and still hold no impact in terms of clinical management. The fifth tier considers the efficacy in terms of patient clinical outcomes (e.g. morbidity or mortality outcomes). The authors pointed that this level of evidence typically require randomised controlled trials or, alternatively, decision-analytic approaches based on existing empirical data and/or assumptions. Finally, the sixth tier considered the overall efficacy associated with the distribution of resources from a societal perspective.

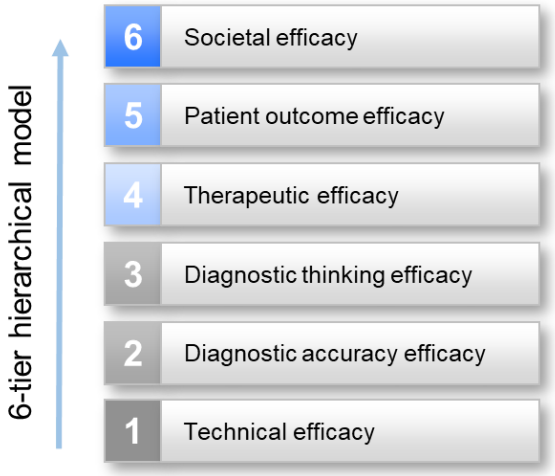


Figure 16. Six level efficacy model [adapted from (Fryback and Thornbury 1991)].

The hierarchic nature of this evaluation framework means that for an imaging test to be efficacious at higher levels, it also needs to be efficacious at lower levels, whilst the opposite is not necessarily true (Fryback and Thornbury 1991). The innovative nature of the model proposed by Fryback and Thornbury (1991) was based on the evaluation of the contribution of medical imaging as a part of a

broader system whose ultimately goal is to improve care for patients as opposed to the traditional concerns about the quality or accuracy of diagnostic tests. Furthermore, the authors proposed the design of before and after tests to estimate the test's impact on the clinician's medical decision algorithm (Fryback and Thornbury 1991).

Subsequent models have added upon the one proposed by Fryback and Thornbury (1991). Kent and Larson (1992) proposed three dimensions of analysis: disease, type of assessment, and the quality of research methods. Whilst the dimension type of assessment is similar to the model proposed by Fryback and Thornbury (1991), the dimension disease describes the clinical condition(s) shown by a diagnostic test and the dimension quality of research methods is classified into four groups (excellent, good, fair or poor) (Kent and Larson 1992). Mackenzie and Dixon (1995) proposed a 5-level framework, very similar to the one defined by Fryback and Thornbury (1991). Other authors, like Silverstein and Boland (1994) and particularly Pearl (1999), further explored the initial model presented by Fryback and Thornbury (1991), providing specific examples on how the different levels of efficacy might be estimated. Similar to previous authors, van der Schouw, Verbeek, and Ruijs (1995) defended a phased approach to diagnostic tests, reserving the inclusion of expensive research clinical trials to influence the decision to actually use or not use the test. However, all these authors proposed specific adjustments to the original model by Fryback and Thornbury (1991) rather than a novel evaluation framework.

2.3.3.2. Houn et al. (2000): 4-phase evaluation framework

Houn et al. (2000), members of the Food and Drug Administration (FDA), proposed a novel framework aimed at providing practical recommendations regarding the evaluation of diagnostic tests. With this aim in mind, the authors explored the rationale of establishing a parallel between the assessment of diagnostic tests and therapeutic methods. These authors proposed a four-phased model for test evaluation similar to the existing framework for therapeutic technologies such as pharmaceutical drugs (Figure 17).

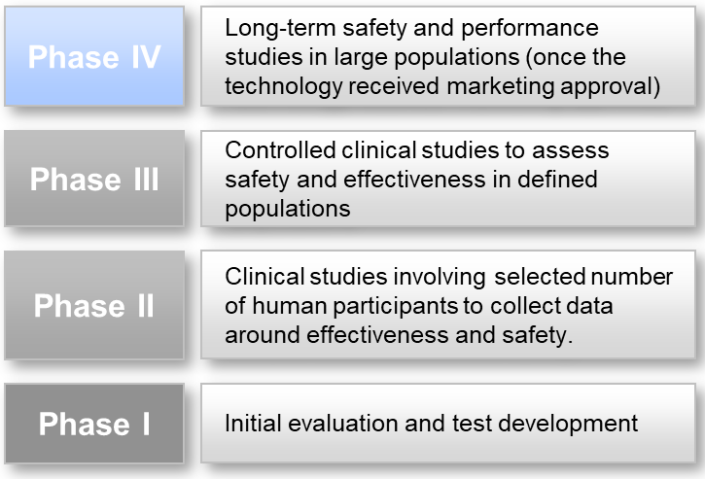


Figure 17. Four level efficacy model [adapted from Houn et al. (2000)].

Despite both aiming to assess the efficacy of diagnostic tests, the models proposed by Fryback and Thornbury (1991) and Houn et al. (2000) constituted two different approaches. Whilst Fryback and Thornbury (1991) proposed a model discerning the different levels from a more basic level, technical accuracy, to the highest level, the implications of accuracy to society, Houn et al. (2000) proposed a research-based framework with different levels of efficacy implicitly embedded in the 4-phase model. Figure 18 illustrates the equivalence between these two types of models.

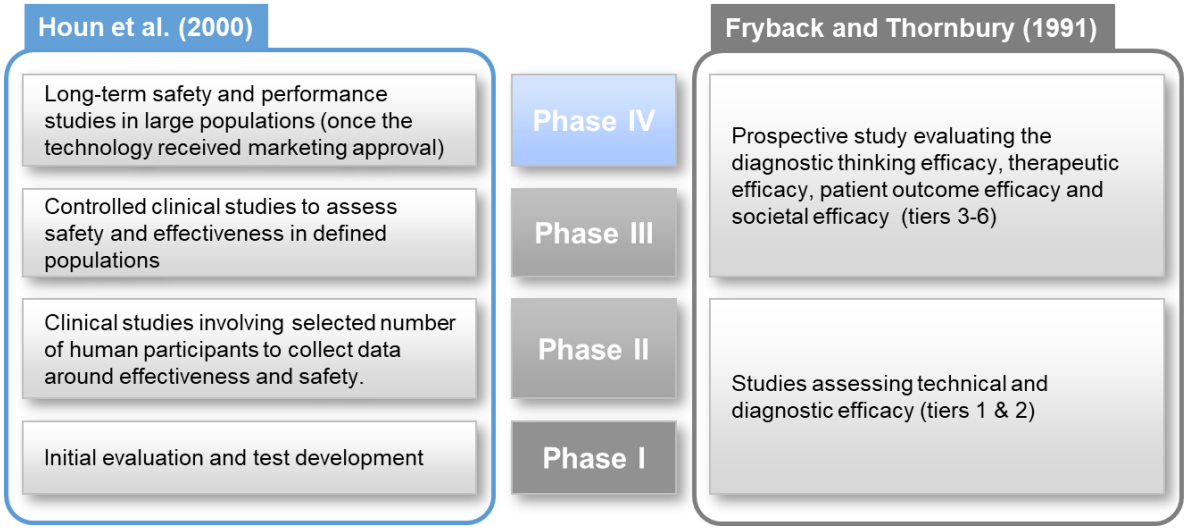


Figure 18. Equivalence in the evaluation of efficacy between the models proposed by Fryback and Thornbury (1991) and Houn et al. (2000).

Despite the conceptual differences, the evaluation models present several common features (Lijmer, Leeflang, and Bossuyt, 2009): (i) start with a phase I, i.e. test development around the technical and diagnostic accuracy of the test; (ii) the application of the diagnostic test to a specific group of individuals with a given clinical disease; (iii) the impact of the test in the clinician's therapeutic decision; (iv) the impact of the test at patient-level and societal outcomes; and (v) the hierarchical nature of the model, i.e. the evaluation of efficacy at higher levels only if in the presence of positive evidence at lower levels.

The systematic literature review by Lijmer, Leeflang, and Bossuyt (2009) considered evidence published up to January 2009. Hence, any model created over the past decade was not included. In order to supplement the information provided by Lijmer, Leeflang, and Bossuyt (2009), a second systematic literature review was performed by the student to update the timeline up to November 2018. The search strategy considered was equal to the one proposed by Lijmer, Leeflang, and Bossuyt (2009) and is detailed in Appendix I. The systematic search of academic publications was conducted in accordance with national guidance (Centre for Reviews and Dissemination 2009) and based on PRISMA guidelines (Higgins and Green 2011) as illustrated in Figure 19. The following databases were considered: Ovid Classic and EMBASE (1990 to 23 November 2018), Ovid MEDLINE(R) In-Process & Other Non-Indexed Citations, Ovid MEDLINE(R) (1990 to 23 November 2018), Cochrane Library NHS Economic Evaluation Database and Cochrane Library CRD Health Technology Assessment Database (1990 to 23 November 2018).

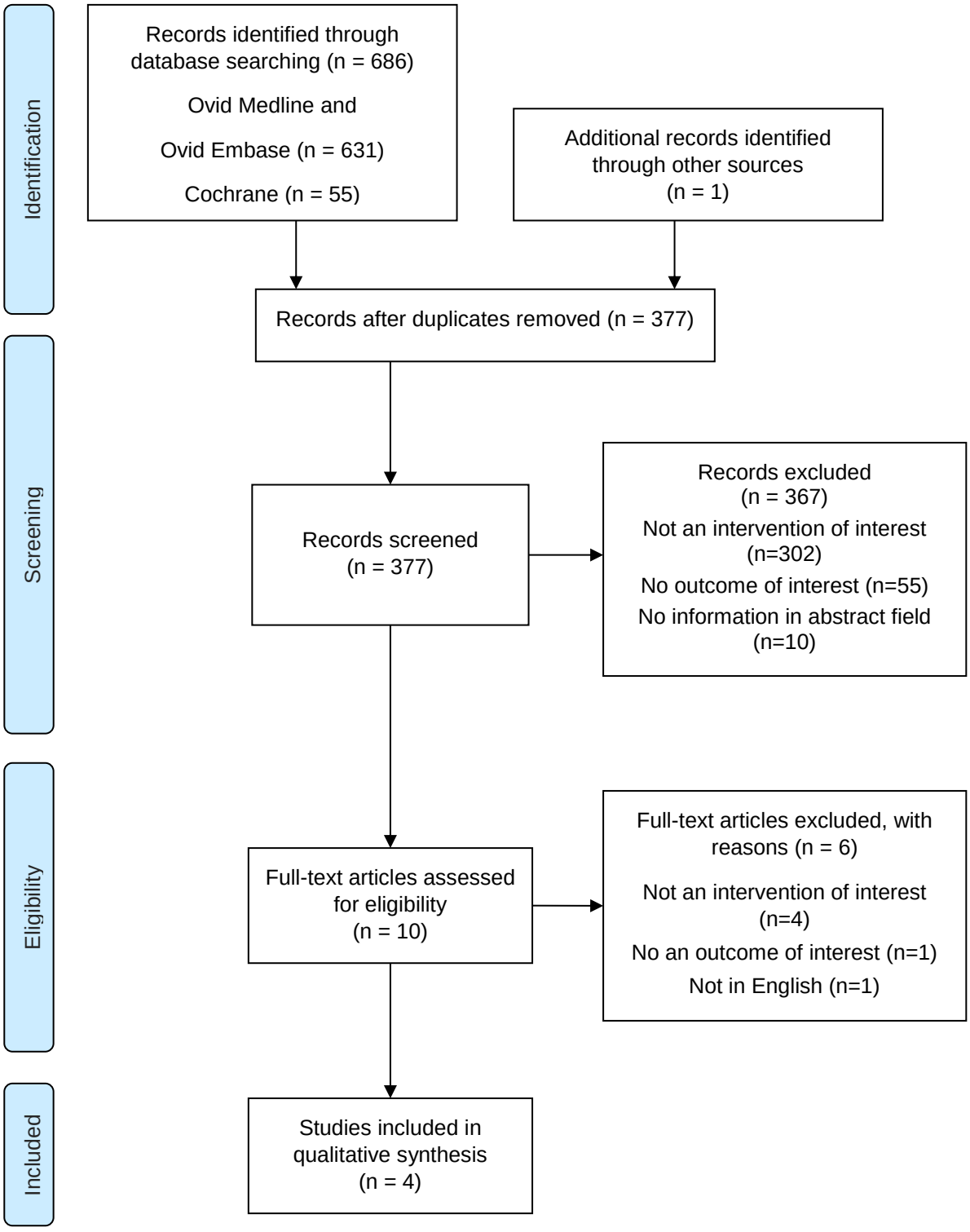
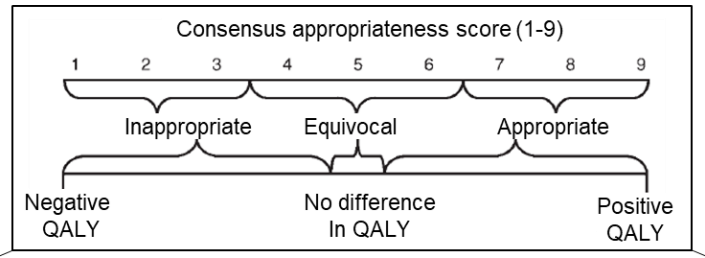


Figure 19. PRISMA flowchart summarising the selection process of relevant studies.

A total of 687 papers were identified, resulting in 377 after removal of duplicates. Out of the 377 papers screened, a total of 10 papers were included for full-text analysis. Six papers were excluded as they did not present an actual evaluation model or framework of medical imaging technologies (n=4), did not consider outcomes of interest (n=1) or were not written in English (n=1). Hence, a total of four papers were analysed and included in the present systematic literature review: Sistrom (2009); Gazelle et al. (2011); Anonychuk et al. (2012); and Frueh and Quinn (2014). The chronological description of the four evaluation frameworks is presented in the following paragraphs, with particular emphasis on Gazelle et al. (2011).

Sistrom (2009) proposed the appropriateness framework. Based on the principle that decision analytic models are more efficient than clinical trials, the appropriateness method considers the utilisation of experts in the context of a set of clinical scenarios. Two groups of experts rate from 1 (very inappropriate) to 9 (most appropriate) the relation between the intervention and the set of clinical scenarios. Then, Delphi rounds further discuss the scenarios where no consensus was reached. The authors proposed a direct link between the 1-9 scale and the appropriateness scale based on the quality-adjusted life years (QALY) approach. Sistrom (2009) proposed, as illustrated in Figure 20, that scores: of 1-3=inappropriate; 4-6=equivocal; and 7-9=appropriate.



Recommendation	Description	Appropriateness score
1. Indicated	Most likely to contribute to clinical diagnosis and management	Appropriate (score 7-9)
2. Specialised investigation	Complex, time-consuming, or resource-intensive investigations performed after discussion with radiologist in the context of locally agreed protocols	Equivocal (score 6)
3. Not indicated initially	Experience shows that clinical problem usually resolves with time, and panel suggests deferring study and only performing if symptoms and/or signs persist and are important to warrant action	Equivocal (score 5)
4. Indicated only in specific circumstances	Non-routine studies that will be performed only if physician provides cogent reasons or if radiologist believes examination represents an appropriate way of furthering diagnosis and management	Equivocal (score 4)
5. Not indicated	Supposed rationale for investigation is untenable	Inappropriate (score 1-3)

Figure 20. Schematic proposing a superimposed relation between the appropriateness consensus score and the concept of QALY (based on Sistrom 2009).

In essence, the authors proposed the use of clinical experts to determine the clinical appropriateness and expected benefits from using the medical imaging technology in the context of given clinical scenarios. Furthermore, based on the appropriateness score, a recommendation whether or not to adopt a given medical imaging technology is considered. This evaluation framework has important limitations as it is only focussed on the clinical appropriateness of the technology and seems to fail to recognise its cost implications to the healthcare payer or society overall.

More recent models have aimed to further capture the intrinsic complexity of diagnostic tests, particularly: the multiple and dynamic applications of diagnostic tests (e.g. MRI being used across different clinical conditions); the rapid change associated with diagnostic technologies; inter-reader and machine variability; and the non-clinical impact of diagnostic tests due to their ability to affect patients and clinicians' perceptions and behaviours. One such model was proposed by Gazelle et al. (2011).

2.3.3.3. Gazelle et al. (2011): 6-tier framework with 3 dimensions of analysis

Further expanding on the original model by Fryback and Thornbury (1991), Gazelle et al. (2011) developed a novel conceptual framework that combined different elements from previous frameworks. The authors suggested that this framework should be used to guide the approval or introduction of new or novel diagnostic technologies (Gazelle et al. 2011). Gazelle and colleagues (2011) based their evaluation framework on three key features: (i) size of the at-risk population, i.e. the number of people that might benefit from the diagnostic intervention; (ii) the anticipated clinical impact, or, the potential net benefits in terms of health outcomes compared to existing alternatives and standard care; and (iii) the potential economic impact, i.e. including cost-effectiveness concepts and the potential financial and budget implications surrounding the adoption of a diagnostic test (Figure 21).

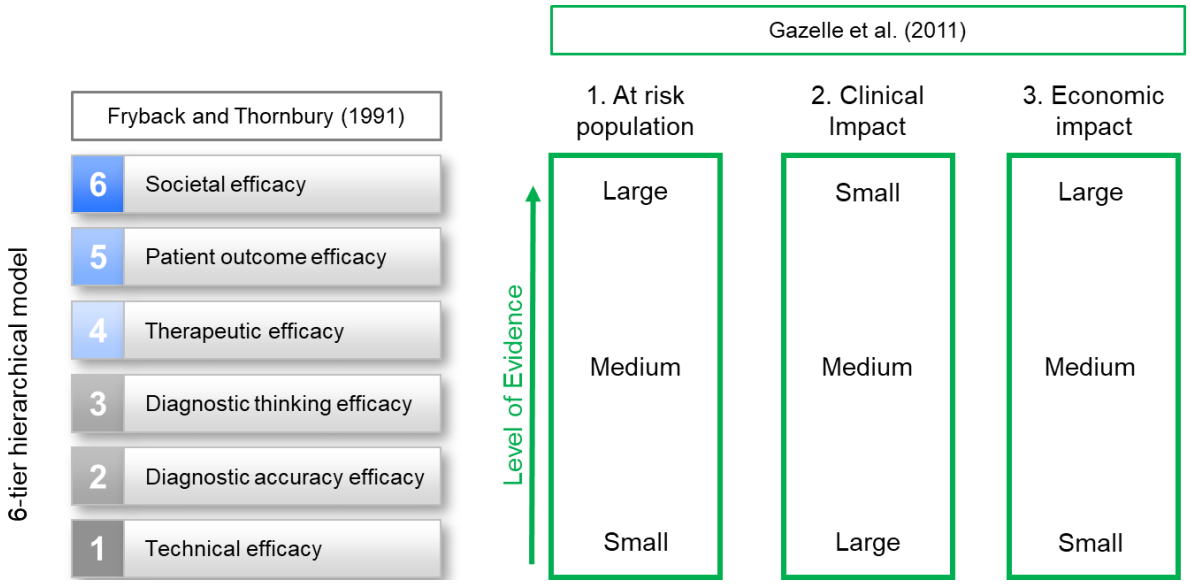


Figure 21. Dimensions of analysis and level of evidence in the model proposed by Gazelle et al. (2011).

In order to enhance its usability, Gazelle et al. (2011) proposed a link between the three features above mentioned and the level of evidence required to evaluate a diagnostic test. As illustrated in Figure 21, if a large population could benefit from a diagnostic test, a higher level of evidence is required in order to evaluate the test's efficacy [equivalent to the higher tiers of the model proposed by Fryback and Thornbury (1991)]. However, if a small number of patients is at risk then a lower level of evidence is indicated [equivalent to the lower tiers of the model proposed by Fryback and Thornbury (1991)]. Similarly, if the intervention is anticipated to have a small clinical effect or hold a potentially large economic impact, there is a need for high-level evidence. In other words, if the potential economic impact of the diagnostic test is large, then a higher level of empirical evidence is required to better understand the potential health economic and budget implications of implementing the intervention.

Following on Gazelle and colleagues' work, Lee, Neumann, and Rizzo (2010) suggested the creation of three dimensions of analysis: medical value, i.e. the test's ability to inform a clinical decision; planning value, i.e. the test's ability to inform patients and clinicians; and 'psychic' value, i.e. the test's ability to affect patient satisfaction or behaviours. Anonychuk et al. (2012) also considered the holistic impact of diagnostic tests across the continuum of care and suggested three dimensions of analysis based on the test's role in the clinical pathway (screening, diagnosis, treatment selection, prognosis and monitoring). These three dimensions are illustrated in Figure 22: (i) optimisation of operational efficiencies; (ii) optimisation of patient management; and (iii) influence on patient behaviour and other effects.

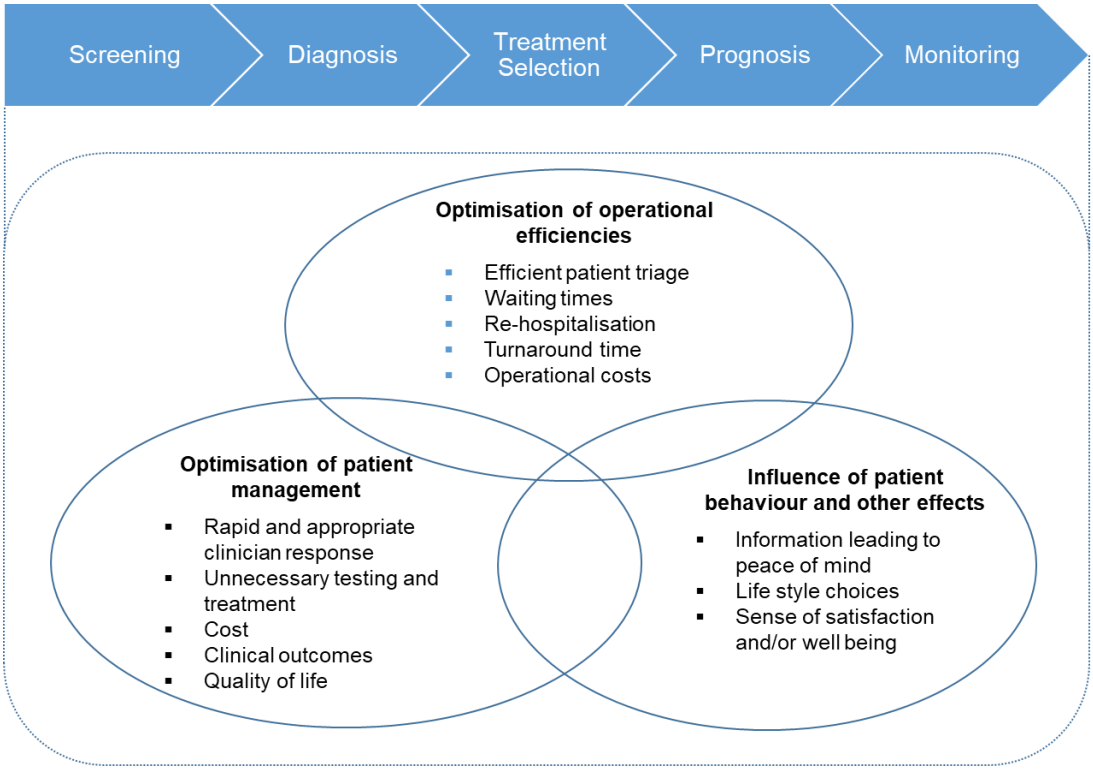


Figure 22. Three dimensions of analysis in the model proposed by Anonychuk et al. (2012).

Compared to Gazelle et al. (2011), the main innovative feature of the evaluation framework suggested by Lee, Neumann, and Rizzo (2010) and Anonychuk et al. (2012) was to explicitly consider the human component in terms of patient behaviours and well-being in the evaluation of diagnostic tests. Intangible human benefits can impact the overall healthcare system. As an example, a diagnostic test can provide reassurance to patients, leading to behavioural changes that ultimately affect the entire clinical pathway and healthcare resource utilisation.

Building upon previous evidence, Frueh and Quinn (2014) proposed a six-part framework, based on the following set of six questions rather than hierarchical levels [opposed to Fryback and Thornbury (1991)]:

1. Who should be tested and under what circumstances?
2. What does the test tell us, that we did not know without it?
3. Can we act on the information provided by the test?
4. Does the outcome change in a way we find value in, relative to the outcome(s) obtained without the test?
5. Will we act on the information provided by the test?
6. If the test is to be employed, can we afford it?

This framework, although developed for molecular diagnostic technologies rather than the medical imaging field, aimed to fill the gap between more conceptual frameworks [e.g. Fryback and Thornbury (1991) or Houn et al. (2000)] and the actual decision and reimbursement process associated with the adoption of new or novel diagnostic tests. More concretely, the authors pointed out that although there is an increasingly need for clinical and economic evidence (e.g. cost-effectiveness), there is little guidance as to how to pragmatically apply this in the context of complex, real-world healthcare delivery settings (Frueh and Quinn 2014). Conversely, this framework's drawback seems to be due to its lack of structure and ability to identify the level of evidence required to evaluate different medical imaging modalities in different contexts.

2.3.3.4. National Institute for Health and Clinical Excellence (2011)

The National Institute for Health and Clinical Excellence (NICE) is the organisation responsible for evaluating the introduction of new or novel technologies in the NHS. Among these, different imaging technologies are considered. This subsection describes the evaluation framework proposed by NICE and the practical aspects associated with the introduction of technologies in the NHS.

Two different evaluation programmes are considered in the evaluation of medical imaging technologies: the Medical Technologies Evaluation Programme (MTEP); and the Diagnostics Assessment Programme (DAP). As illustrated in Figure 23, the two programmes differ in the value proposition associated with the diagnostic test. Whilst in the MTEP, the technology must have an equivalent or superior clinical performance and no increase in costs, in the DAP the technology must

show increased health benefits but at a higher cost or cost savings at expense of reductions in health benefits. As suggested by Gazelle et al. (2011), NICE considers the implicit evaluation of two dimensions of analysis: the clinical and the economic impact of the intervention. Hence, technologies with potential impact in the clinical dimension with no economic impact are routed through the MTEP. Conversely, diagnostic technologies with potential to significantly impact the clinical and economic dimensions are routed through the DAP.

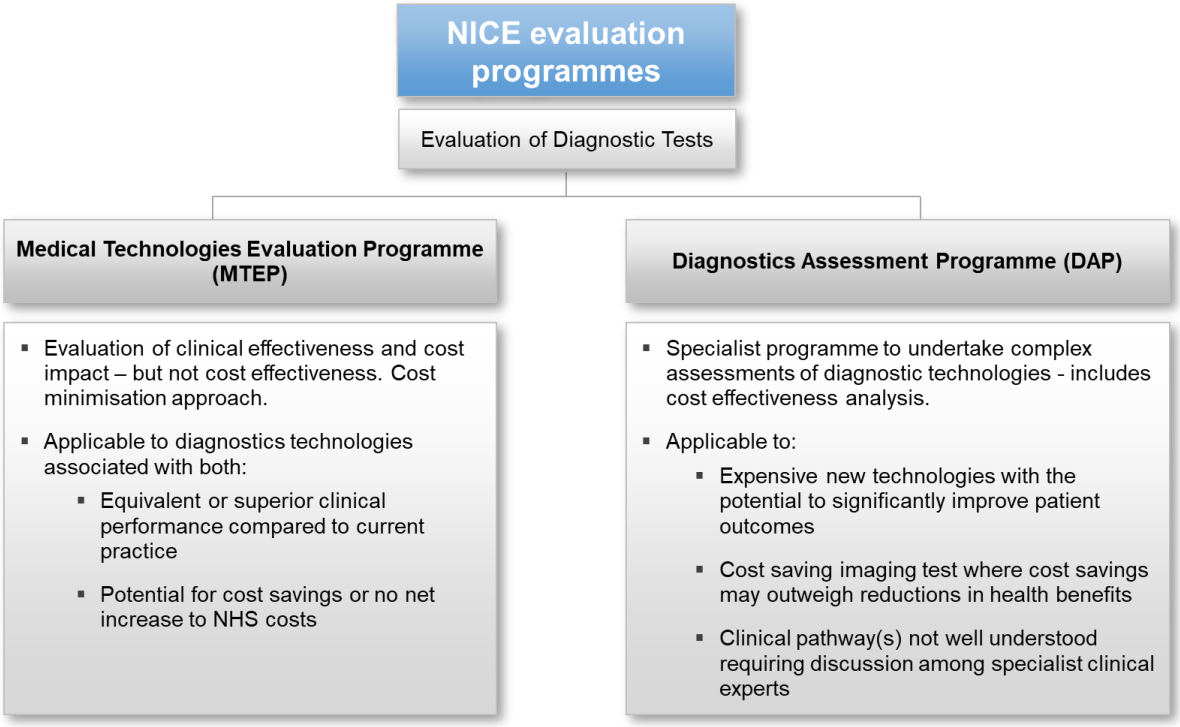


Figure 23. Two evaluation programmes of diagnostic tests considered by NICE (Crabb, 2011).

The existence of two programmes with different methodologies to evaluate diagnostic technologies reflects not only the potential impact of such technologies but also the complexity associated with their evaluation. For illustration purposes, a well-known technology applied to a new clinical pathway that is likely to be associated with clinical improvements at no extra cost would be evaluated using the MTEP methodology. This methodology is essentially based on a cost-consequence approach (or cost-minimisation if outcomes are assumed to be equivalent) (NICE 2011d). However, a novel imaging modality or a diagnostic test used in a disruptive way is usually associated with a higher degree of uncertainty, both in terms of its clinical and economic impact. This technology would be evaluated under the DAP methodology, based on cost-effectiveness analyses (NICE, 2011).

Both evaluation programmes require a systematic literature review of clinical and economic evidence. Furthermore, given the limited availability and variability of evidence, NICE does not have specific evidence thresholds for the evaluation of diagnostic tests, but rather relies on the existing evidence so

as not to delay the introduction of such technology (Crabb, 2011). Thus, compared to pharmaceutical products, NICE does not require evidence from randomised trials and instead assesses the level of evidence on a case by case basis.

2.3.3.5. Summary of evaluation frameworks

Different evaluation frameworks for diagnostic tests have been proposed since the 6-tier model originally suggested by Fryback and Thornbury (1991). Recent models have tried to incorporate different features to take into consideration usability in the context of real-world decision making processes. Among these features, two were dominant: (i) the hierarchical nature of the evaluation framework proposed; and (ii) the level of evidence necessary to inform a decision regarding the use of diagnostic tests.

Most evaluation frameworks rely on a hierarchical structure (Fryback and Thornbury 1991; Houn et al. 2000; Gazelle et al. 2011). Lower tiers of the evaluation frameworks relate to intrinsic characteristics of the diagnostic tests, such as sensitivity and specificity, whilst higher tiers evaluate the actual clinical and economic impact of the intervention in the context of real-world clinical pathways. Diagnostic tests that fulfil the criterion of higher tiers of the evaluation framework inherently meet the criteria of lower tiers of the model, i.e. a diagnostic test that presents favourable clinical and economic evidence presents higher levels of accuracy for a given clinical condition. The opposite might not be the case, i.e. diagnostic tests that fulfil the lower tiers of the evaluation framework might not be associated with improved clinical and economic benefits (higher tiers of the model). The hierarchical nature of most evaluation frameworks is designed to provide a logical approach to a complex problem.

The level of evidence necessary for policy makers to adopt/reject specific diagnostic tests in real-world clinical contexts is the other key feature of some of the proposed evaluation frameworks. Recent models tend to focus more on this aspect. Gazelle et al. (2011) were among the first authors to clearly identify the need for evidence thresholds for different types of interventions depending on their impact on different dimensions of analysis. For instance, taking into consideration the rapid development of technologies in the diagnostic field, if an intervention is anticipated to have large clinical benefits for a small number of patients, thus resulting in a limited economic impact, its adoption should not be delayed to gain evidence from a high-quality study (e.g. randomised controlled trial). Moreover, regardless of the population at risk and the intervention's anticipated impact, the evaluation of the economic impact is relevant even in small populations as the use of interventions that are not cost-effective is not considered to be value for money. This pragmatic approach poses an important question, how to establish the threshold of evidence for the adoption of different diagnostic technologies. Is it necessary to use evidence from randomised controlled trials and observational studies or does evidence from economic modelling studies suffice? And if so, under which circumstances? To better understand this issue, it is relevant to investigate the existing economic

evidence around the utilisation of medical imaging. With this in mind, a literature search identified the following two systematic literature reviews:

- Otero, Hansel J., Frank J. Rybicki, Dan Greenberg, and Peter J. Neumann. 2008. “Twenty Years of Cost-Effectiveness Analysis in Medical Imaging: Are We Improving?” Radiology 249 (3): 917–25. This systematic literature review considered the analysis of any published cost-effectiveness analysis of diagnostic tests for a period of 20 years (1985-2005).
- Zhou, Alice, David M. Yousem, and Matthew D. Alvin. 2018. “Cost-Effectiveness Analysis in Radiology: A Systematic Review.” Journal of the American College of Radiology 15 (11): 1536–46. Consistent with Otero et al. (2008), this systematic literature review included any cost-effectiveness analysis study for a period of five years (2013-2017).

These two systematic literature reviews cover a large period of time (25 years) and the second review was published in November (2018), referring to the last five years of evidence (2013-2017). For these two reasons, it was not considered relevant for the student to perform another review but to examine the existing evidence, particularly by Zhou, Yousem, and Alvin (2018).

2.3.4. Economic evidence

This subsection summarises the methodology regarding the evaluation of medical imaging technologies. More than the economic results themselves, which are arguably context and intervention-specific, the aim here is to evaluate different features associated with the published evidence, and assess the potential impact of evaluation frameworks on the actual assessment of real-world interventions.

Otero et al. (2008) performed a systematic literature review of cost-effectiveness analyses in the medical imaging field published between 1985 and 2005. The authors searched several databases: MEDLINE, HealthStar, CancerLit, Current Contents, EconLit and Health Economic Evaluation Databases using broad text keywords such as “QALY”, “quality-adjusted”, “cost-utility” (Otero et al. 2008). Despite the increase in false positives, the search strategy was designed to capture all relevant medical imaging related cost-effectiveness studies with QALYs as the measure of effect. A total of 1,310 articles were screened, 111 of which were included for full-text analysis (Otero et al. 2008).

About three quarter of the evidence was generated in the US, with only 9 (8.1%) papers being originated in the UK (Otero et al. 2008). A total of 86 (77.5%) papers considered the evaluation of medical imaging as a diagnostic procedure, with the remaining 25 (22.5%) manuscripts reporting its use as an interventional procedure. Other key characteristics are summarised in Table 2.

Table 2. Summary characteristics of the studies analysed by Otero et al. (2008).

Publishing Year	Number of papers	%
2000-2005	62	55.9%
1995-1999	41	36.9%
1990-1994	6	5.4%
1985-1989	2	1.8%
Quality of papers	Mean quality rating (Likert scale 0-7)	
2000-2005	4.03	
1995-1999	4.34	
1990-1994	4.00	
1985-1989	4.00	
Modality of imaging	Number of papers	%
Ultrasound	39	35.1%
Angiography	35	31.5%
MRI	25	22.5%
CT	22	19.8%
Conventional radiography	10	9.0%
PET or combined PET/CT	8	7.2%
Other	11	10.0%
Clinical condition in study	Number of papers	%
Peripheral vascular disease (non-cerebral, non-cardiac disease)	25	22.5%
Cancer	20	18.0%
Cerebrovascular disease	15	13.5%
Ischaemic heart disease	15	13.5%
Musculoskeletal and rheumatologic diseases	9	8.1%
Other	27	24.3%
Perspective of analysis	Number of papers	%
Payer	76	68.5%
Societal	31	27.9%
Hospital	4	3.6%

There was an increase in the volume of papers published in recent years. In fact, the second decade in analysis (1995-2005) produced over 90% of the total evidence generated in the two decades analysed. Nevertheless, the overall quality of the evidence generated - measured subjectively based

on a 1 (low) to 7 (high) Likert scale - did not seem to have improved over this period of time. This suggests that, although health economics is increasingly relevant to the medical imaging field, this trend has not been followed up by an improvement in the overall quality of the economic evidence. Otero et al. (2008) pointed out that this seemed to be due to concerns about the lack of agreement in the methodologies used, and potential biases due to data selection and/or model development opacity. This finding suggests that, despite the creation and development of several evaluation frameworks (previously described) which took place between 1985 and 2005, their respective impact in the generation of high-quality economic evidence was limited.

Ultrasound and angiography were the two modalities of medical imaging with the highest number of economic evaluations with 39 (35.1%) and 35 papers (31.5%), respectively. The use of advanced imaging modalities, such as CT and MRI, were the third and fourth imaging modalities more commonly evaluated. With regards to the clinical condition, peripheral vascular disease (n=25, 22.5%), cancer (n=20, 18.0%) - particularly lung cancer - cerebrovascular disease (n=15, 13.5%) and ischaemic heart disease (n=15, 8.1%) were the most common. In relation to the study design, only 3 (2.7%) papers considered economic data retrieved from randomised clinical trials.

The second systematic literature review, performed by Zhou, Yousem, and Alvin (2018), analysed the economic literature regarding any cost-effectiveness or cost-utility analyses in the imaging field between the years 2013 and 2017. The study's aim was to evaluate the methodological variation in the economic assessment of medical imaging interventions and explore impact on the overall results. In order to achieve this aim, the authors opted for a broad search criteria, designed to enhance the search's sensitivity (i.e. capture all relevant papers), despite the increase in the number of non-relevant hits. The databases MEDLINE, EconLit and Tufts CEA were queried using the following search criteria: (((cost-effectiveness) OR (cost-utility)) AND (imaging OR radiology). A total of 2,574 non-duplicate records were screened and a total of 240 full-text papers were assessed. Out of these, the authors limited the search to cost-utility analyses (i.e. cost per QALY), retrieving 80 articles to be included in the systematic literature review (Zhou, Yousem, and Alvin 2018).

The authors rated the included 80 full-text papers using the Consolidated Health Economic Evaluation Reporting Standards (CHEERS) checklist (Husereau et al. 2013), designed to assess cost-utility analyses. This checklist includes 24 items against which each paper was evaluated. The economic evidence highlighted a wide variation in items, particularly cost estimates, outcome measurement, and the analytical and statistical methods used (Zhou, Yousem, and Alvin 2018). The authors pointed out a lack of transparency surrounding the methods used with potential to impact the results of the health economic evaluations (Zhou, Yousem, and Alvin 2018). According to Zhou, Yousem, and Alvin (2018), the clear dissemination of transparent recommendations and evaluation frameworks for cost-effectiveness analyses in the radiology field is essential to support policy makers in their decision to whether or not to include a medical imaging technology as part of standard care. The 80 papers included in the systematic review were analysed and grouped by: year of publishing; type of imaging;

clinical condition; perspective of analysis; and study design. This evidence is summarised in Table 3 (Zhou, Yousem, and Alvin 2018).

With regards to the publishing year, there was no noticeable change over the five year period, with an average of 16 papers per year. This in contrast to the increase in evidence reported by Otero and colleagues (2008), where over 90% of evidence was generated in the second decade in analysis.

The use of advanced imaging modalities such as CT (n=45, 56.3%) or MRI (n=28, 35.0%) was considered in over 90% of papers. Again, this was a noticeable difference from the evidence reported by Otero et al. (2008), where the evaluation of conventional imaging was more prevalent than advanced imaging modalities. Nevertheless, the findings reported by Zhou, Yousem, and Alvin (2018) were not unexpected as they were consistent with the historical increase in the utilisation rates associated with advanced imaging modalities.

The economic evaluation of medical imaging was used across multiple clinical conditions, with particular relevance to cancer (32%) and cardiovascular diseases (12%). The two cancers more prevalent were breast and lung cancer, accountable for 39% and 33% of all cancer-related economic evaluations.

Regarding the perspective of the economic analyses, 74% of the papers analysed took a health care payer perspective, followed by a societal perspective in 20% of the analyses. This is consistent with the evidence reported by Otero et al. (2008) and reflects the typical use of economic evaluations as a critical supporting tool to health care decision makers to whether adopt or reject a specific intervention.

Eighty-seven percent of the papers were based on probabilistic models (mainly Markov and/or discrete event simulation models) with the remaining 13% of evidence based on observed data from randomised controlled trials (n=4) and prospective cohort studies (n=7). This is similar to the evidence reported by Otero et al. (2008). From the supplementary material, only three actual RCTs were identified (Dekkers et al. 2016; Thom et al. 2014; Agus et al. 2016). Two of these three RCTs were conducted in the NHS (Thom et al. 2014; Agus et al. 2016) and they evaluated the cost-effectiveness of advanced imaging (CT) in patients presenting with stable chest pain. The seven prospective studies included different stroke (Parody et al. 2015), chest pain (Hlatky et al. 2015) and cancer, both breast (Pharoah et al. 2013) and lung cancer (Yang et al. 2017; Pertile et al. 2015; Black et al. 2014; Gómez León et al. 2014). In summary, over 9 out of 10 papers were not based on observed data but rather decision analytical models or uncontrolled observed data. This situation is explained by multiple factors, particularly: (i) the time it takes to generate evidence from real-world studies; (ii) the rapid technological developments in the medical imaging field; and (iii) the lack of a common and transparent framework to evaluate medical imaging.

Table 3. Summary characteristics of the studies analysed by Zhou, Yousem, and Alvin (2018).

Publishing Year	Number of papers	%
2017	17	21.3%
2016	14	17.5%
2015	18	22.5%
2014	16	20.0%
2013	15	18.8%
Modality of imaging	Number of papers	%
CT	45	56.3%
MRI	28	35.0%
Ultrasound	11	13.8%
Nuclear Medicine	6	7.5%
Radiography	3	3.8%
DEXA	2	2.5%
Digital mammography	13	16.3%
Tomosynthesis	2	2.5%
Clinical decision rule	5	6.3%
Clinical condition in study	Number of papers	%
Cancer	33	41.3%
Cardiovascular disease	12	15.0%
Intracerebral haemorrhage	9	11.3%
Bone imaging	7	8.8%
Blunt cerebrovascular trauma	4	5.0%
Varied diagnoses	15	18.8%
Perspective of analysis	Number of papers	%
Payer	56	70.0%
Societal	15	18.8%
Hospital	2	2.5%
Societal and payer	3	3.8%
Study design	Number of papers	%
Probabilistic model	71	88.8%
Non-model	11	13.8%
Randomised Controlled Trial	4	
Prospective Cohort	7	

In summary, although there have been iterative improvements in the economic evaluation of medical imaging technologies, limited consideration has been given to the potential impact of using evidence generated from modelling evidence as opposed to real-world observed data. This poses two questions:

- Are the economic findings of a new or novel medical imaging intervention dependent on the study design (i.e. economic models vs non-model approaches)?
- If yes to the above, would this difference in economic findings result in different adoption scenarios from the decision maker (e.g. adopt, reject, adopt for a limited cohort of patients)?

The responses to these two questions remain unclear given the limited number of comparative studies. Hence, the present thesis aims to contribute to this debate by comparing *a priori* decision-analytical modelling approaches to a real-world randomised clinical trial and prospective cohort studies concerning the utilisation of advanced imaging (discussed in chapter 6). The medical imaging interventions considered include the use of advanced imaging (either CT or MRI) in the context of three specific NHS clinical pathways: (i) suspected scaphoid fracture (chapter 3); (ii) chronic headache (chapter 4); and (iii) suspected colorectal cancer (chapter 5).

2.4. From research to clinical practice

Faced with ever increasing healthcare research, decision makers struggle to keep up with the rapidly evolving evidence (Wensing and Grol 2019). This poses considerable strain on decision makers to ensure the rapid uptake of high-value clinical procedures, technologies and organisational models into routine clinical practice and, at the same time, stop the use of interventions that no longer represent value for money (Wensing and Grol 2019). Many interventions with favourable research evidence have failed to translate into meaningful improvements, with some estimates indicating that up to two-thirds of initiatives to implement change fail (Damschroder et al. 2009). The understanding of healthcare specific, and general barriers to implementation, are essential to ensure that research evidence is successfully translated to clinical practice.

2.4.1. Barriers to implementation

Silva (2015) conducted a systematic review in twelve bibliographic databases using broad implementation terms (e.g. implementation, change, adoption, feasibility) to assess the evidence concerning implementation initiatives in the context of the NHS. A total of 73 articles met the inclusion criteria, with 53% focused on hospital initiatives, 16% on primary care and 30% on combined primary and secondary care services. Regardless of the different nomenclatures used in literature, evidence demonstrated that most barriers to implementation resulted from: the intervention itself; characteristics of the individuals involved; organisational factors; and contextual or environmental factors.

Figure 24 illustrates a diagram (from micro to macro level) with the five most common barriers to implementation in healthcare (Grol and Grimshaw 2003; Rubio-Valera et al. 2014; Silva 2015; Fischer et al. 2016; Sommerbakk et al. 2016).

First, some barriers are due to the intervention itself, which may be perceived as too difficult to use, incompatible with usual routines, too costly or associated with too little or too much evidence.

Second, characteristics of the individuals may act as barriers to implementation. These range from natural resistance to change to individual beliefs and past experiences (e.g. change as a cost cutting exercise), as well as more intrinsic characteristics such as skills, self-confidence and motivation (lack of motivation as a barrier to implementation).

Third, interpersonal relationships can also act as barriers to implementation. Health care delivery is characterised by the interaction of multidisciplinary teams and any barriers to communication can seriously hinder implementation initiatives. The lack of relationships (formal or informal) between key individuals and the lack of engagement of opinion leaders are two examples of interpersonal barriers.

Fourth, organisational barriers include adverse organisational culture, lack of senior engagement, fragmented systems (e.g. IT systems) and the lack of resources, both in terms of staff, and time availability dedicated to the implementation initiative.

Fifth, the wider involvement of the public and society can generate important barriers. Particularly relevant are the regulatory context (e.g. guidelines, national targets, financial incentives), the NHS setting (e.g. existing culture, fragmented healthcare provision) and patients' and societal expectations. The next subsection assesses how existing implementation models can deal with these barriers.

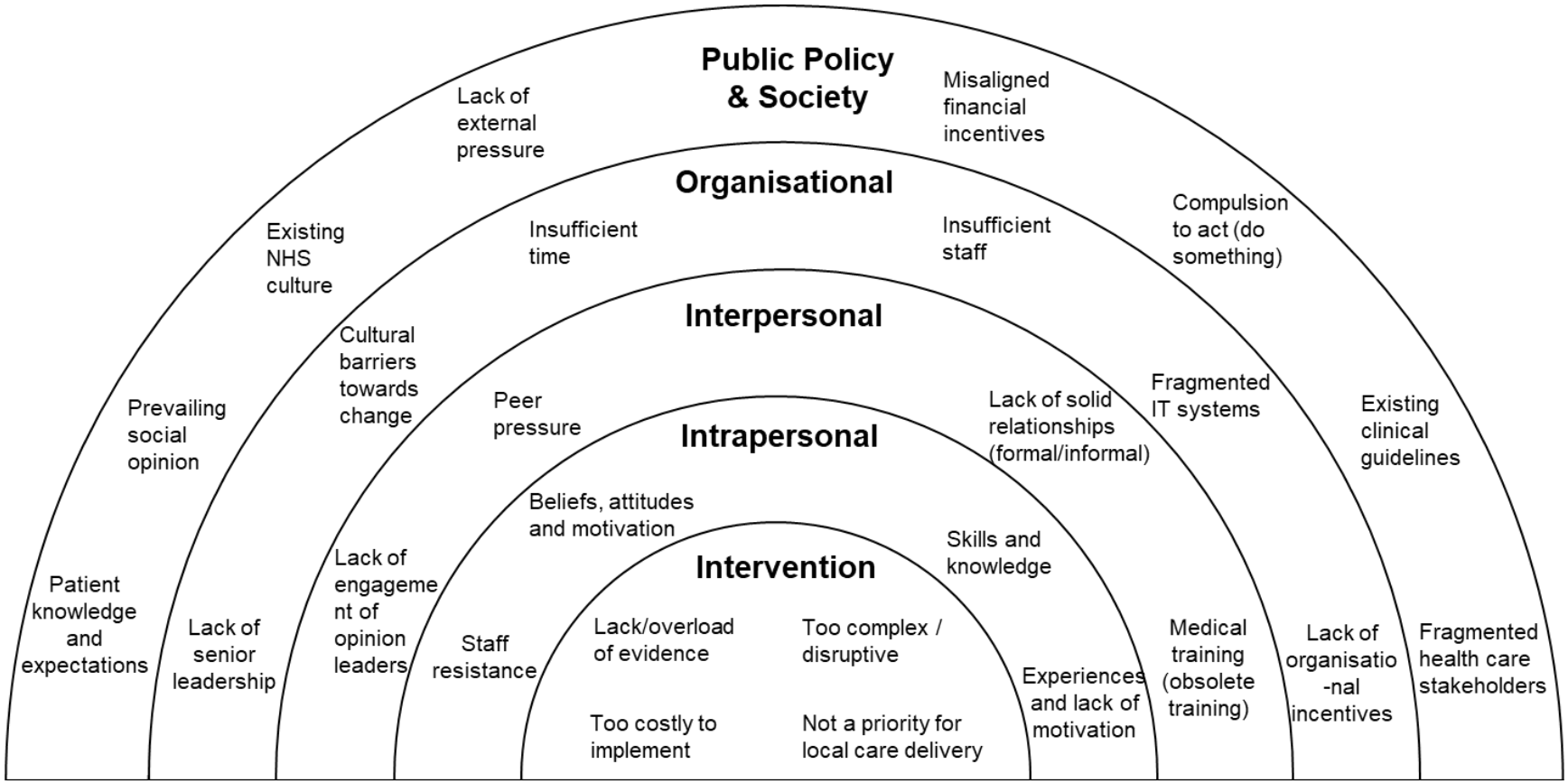


Figure 24. Five domains of barriers or challenges to implementation projects in healthcare (based on Rubio-Valera et al. (2014) and adapted with information from Grol and Grimshaw 2003, Silva 2015, Fischer et al. 2016 and Sommerbakk et al. 2016)

2.4.2. Implementation models

The field that studies the incorporation of healthcare research into clinical practice is labelled using different terms, particularly translational research, dissemination and implementation research and knowledge translation (Titler 2018; Wensing and Grol 2019). All these terms describe the process of applying evidence to practice and, although with slight differences, they seem to overlap and to be used interchangeably in the literature (Wensing and Grol 2019). These terms also seem to be more geographically driven rather than due to substantial conceptual differences. For instance, the concept dissemination and implementation research is often used in Europe and knowledge translation in the USA (Wallace 2013).

Translational research refers to the application of basic research evidence in clinical practice and is commonly divided into multiple stages (T0 to T5). Implementation research focuses on the delivery of interventions to promote the uptake and the use of research to impact patient outcomes (Titler 2018). A detailed comparison of the two concepts showed that stage T3 of translational research which focuses on translation of evidence to clinical practice also includes any type of dissemination and implementation research (Zoellner and Porter 2017). Although the use of different terms can be confusing and might affect the development of the field, all concepts possess a common denominator that is to promote the utilisation of research evidence into clinical practice. For the purpose of this thesis, rather than discussing the different concepts, the aim was to review published evidence and create new evidence in order to facilitate the uptake of the innovative use of advance imaging in the context of three real-world clinical pathways.

Nilsen (2015) extensively evaluated existing implementation theories and frameworks and identified three common aims amongst all models. First, models described or supported the process of translating research into practice. Second, models aimed to understand and/or explain what variables influenced the outcomes of the implementation project. Third, models aimed to evaluate the level of success of the implementation project. Based on the second feature above mentioned, i.e. the model's ability to explain what influenced the implementation outcomes, the author grouped the models or frameworks (terms used interchangeably in most literature) using five categories as per Table 4 (Nilsen 2015).

Table 4. Five categories of models used in implementation research (Nilsen 2015).

Category	Description	Aim(s)
Process models	Models that use specific steps (stages or phases) in the process of translating research into clinical practice. The aim of process models is to provide practical guidance in the planning and execution of implementation initiatives and/or strategies to facilitate implementation. **Example:** The Academic Center for Evidence-Based Practice (ACE) Star Model of Knowledge Transformation.	To describe and/or guide the process of translating research into practice
Determinant frameworks	Models that use specific types (classes or domains) of determinants which act as barriers and enablers (independent variables) that impact implementation outcomes (dependent variables). These models aim to understand and/or predict outcomes prior to implementation. **Example:** Promoting Action on Research Implementation in Health Services (PARIHS) model; and Consolidated Framework for Implementation Research (CFIR).	To understand and/or explain what influences implementation outcomes.
Classic theories	Models that originated from the fields of psychology, sociology and organisational theory and can be applied to provide understanding and/or explanation of aspects of implementation. **Example:** Theory of diffusion.	
Implementation theories	Models developed by implementation researchers to provide understanding and/or explanation of aspects of implementation. **Example:** Organisational readiness.	
Evaluation frameworks	Models that specify aspects of implementation that could be evaluated to determine implementation success. **Example:** RE-AIM framework.	To evaluate implementation.

Given their prominence in the published literature, three models (RE-AIM, PARIHS and CFIR) are described below and chronologically summarised in Table 5.

Table 5. Key features of the three implementation models evaluated (Zoellner and Porter 2017).

Model	Year developed	Key features
RE-AIM	1999	Reach; effectiveness; adoption; implementation; maintenance.
PARIHS	2004	Evidence; context; facilitation.
CFIR	2009	Intervention characteristics; inner setting; outer setting; characteristics of individuals; and process of implementation.

2.4.2.1. RE-AIM model

The RE-AIM framework, originally developed in 1999, aimed to plan and evaluate the implementation of public health initiatives (Glasgow et al. 2019). This framework was based on five dimensions: reach, effectiveness, adoption, implementation and maintenance (Zoellner and Porter 2017; Zoellner et al. 2015; Glasgow et al. 2019). Since its inception, this framework has been widely used. The authors revised the model in 2019 and added key questions associated with each dimension (Table 6).

Table 6. Description of the five dimensions considered in the RE-AIM framework (Glasgow et al. 2019).

Dimension	Definition	Key question
Reach	The number and proportion of the target population that participates in the innovation and its representativeness of the overall population.	“How do I reach the targeted population with the intervention?”
Effectiveness	The impact of an intervention on important outcomes, including quality of life and economic outcomes.	“How do I know my intervention is effective?”
Adoption	The number, proportion and representativeness of: (i) settings; and (ii) people who deliver the intervention.	“How do I develop organisational support to deliver my intervention?”
Implementation	The intervention agents’ fidelity to the implementation initiative, including its duration, frequency and overall cost.	“How do I ensure the intervention is delivered properly?”
Maintenance	The intervention’s ability to maintain its effects at least 6 months following the implementation initiative.	“How do I incorporate the intervention so that is delivered over the long term?”

2.4.2.2. PARIHS model

The PARIHS (Promoting Action on Research Implementation in Health Services) framework, developed in 2004, proposed a way to implement research into clinical practice based on the interaction of three elements: evidence; context; and facilitation (Rycroft-Malone 2004). This framework proposed that the successful implementation of evidence into practice depended not only on the quality of the evidence but also on the context of setting (local, organisational and health system) where the new evidence is introduced as well as the way the evidence is introduced (facilitated into practice). In 2016, the revised PARIHS framework proposed innovation, recipients and context as the three key elements, with facilitation representing the fourth element responsible for the overall alignment and integration of the implementation initiative (Figure 25) (Harvey and Kitson 2015). The 'recipient' element was introduced in the framework following feedback that the original PARIHS framework failed to acknowledge the individuals or groups involved in the implementation (Harvey and Kitson 2015). Both PARIHS frameworks were unique as they introduced facilitation as pivotal in the implementation of any healthcare initiative (Sudsawad 2007).

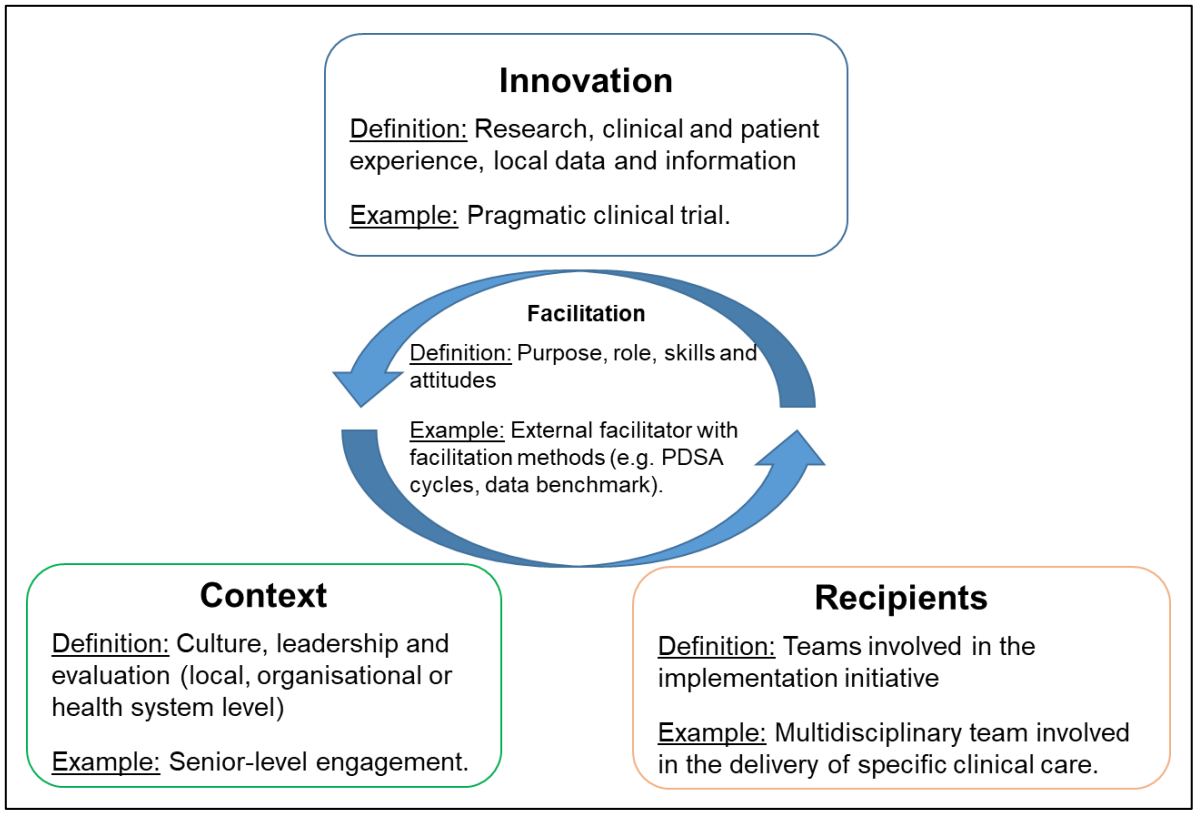


Figure 25. Description of the four elements considered in the revised PARIHS model (adapted from Harvey and Kitson 2015).

2.4.2.3. CRIF model

The Consolidated Framework for Implementation Research (CFIR) embodied a conceptual framework, initially developed in 2009 (Damschroder et al. 2009), based on nineteen published implementation theories (Zoellner and Porter 2017). This framework took a comprehensive approach based on five domains including a total of 37 constructs (see Table 7). Each construct corresponds to a potential barrier or enabler to implementation. The CRIF framework provided a pragmatic structure that ultimately aimed to help individuals and organisations to successfully complete implementation initiatives.

Given the completeness of the CRIF framework, with a total of 37 detailed barriers to implementation, the PhD student decided to use this model to support the NHS implementation of all three advanced imaging initiatives considered in the present thesis. This will be discussed in more detail in Chapter 7.

Table 7. Description of five domains and 37 constructs comprising the CRIF framework [based on Damschroder et al. (2009) and Zoellner et al. (2015)].

Domain	Constructs considered and respective definition
Intervention characteristics	**Intervention source:** internally vs externally developed idea. The legitimacy of the source may influence the success of the implementation (e.g. internal ideas might have more legitimacy compared to external ideas – e.g. consulting company).
	Evidence strength and quality: stakeholders' perception of the quality and validity of the evidence supporting the intervention might impact its implementation. Preferred sources of evidence are published literature and guidelines as opposed to anecdotal stories.
	Relative advantage: stakeholder's perception of the advantage of implementing the intervention versus an alternative solution. In healthcare, it typically refers to the advantage of implementing the intervention as opposed to doing nothing (i.e. maintain standard care).
	Adaptability: the ability to adapt or tailor the intervention to address local needs. The balance between the need to base the implementation on high-quality evidence while addressing specific local challenges.
	Trialability: the ability to test the intervention and, if needed, reverse it. Local pilots or plan-do-study-act (PDSA) cycles are commonly used in healthcare.
	Complexity: stakeholder's perception regarding the difficult of implementation. Disruptive interventions typically require more reorientation and implementation efforts as they reflect a clear departure from the *status quo*. Complexity also increases with the number of organisational units (teams, clinics, departments) or types of people (providers, patient, managers).
	Design quality and packaging: stakeholder's perception of the quality of how the intervention is conducted. The better the intervention is presented, the more likely it is for the implementation to be successful.
	Cost: costs of the intervention itself and the implementation costs associated with its adoption in clinical practice.

Outer settings	**Patient needs and resources**: the degree as to which the patient's needs, as well as barriers/facilitators to address those needs, are known to the organisation and prioritised with the right resources. Patient-centred organisations are more likely to implement change effectively.
	Cosmopolitanism: the degree to which the organisation is linked to external organisations. Organisations that promote external networks are more likely to be successful in the implementation of new practices (notion of social capital).
	Peer pressure: competitive pressure to adopt an intervention. The term 'peers' can refer to any outside organisation or individuals within the same organisation.
	External policies and incentives: external strategies (e.g. governmental or regulatory agencies) that can affect the intervention's implementation (e.g. lack of financial incentives).
Inner settings	**Structural characteristics:** the social architecture (the size and level of independence of each group responsible for delivery of care), age, maturity and size of the organisation. The more stable teams are, the more likely they are to successfully implement change.
	Network and communications: the nature and quality of the social networks among individuals and groups/services/ directorates. Relationships between individuals and cohesive informal communication channels may be more important to a successful implementation than attributes from individuals (sense of 'community' or ' teamness').
	Culture: norms and values of the organisation. Organisations with culture that embrace change are more likely to successfully adopt new clinical practices.
	Implementation climate (with 6 subcategories): tension for change (the degree to which stakeholders perceive the situation as intolerable); compatibility (fit between the meaning and values of the intervention and the individual's own values and perceptions); relative priority (individual's perception of the relevance of the implementation to the organisation); organisational incentives and rewards (extrinsic incentives such as performance reviews, promotions, raise in salary and less tangible incentives such as increase in respect); goals and feedback (the extent to which goals are clearly communicated and fed back to individuals); learning climate (the extent to which individuals feel safe and able to challenge existing methods).

	Readiness for implementation (with 3 subcategories); leadership engagement (commitment and involvement of local and senior leaders); available resources (appropriate level of resources dedicated to support implementation, such as training, space and time); access to information and knowledge (ability to access information about the intervention and how to incorporate it into clinical practice).
Characteristics of the individuals involved	**Knowledge and beliefs about the intervention**: organisations are composed of individuals who have their own knowledge and beliefs toward changing behaviours. This construct, along with self-efficacy (below), are the two most common constructs used in models based on classic and implementation theories.
	Self-efficacy: the individual's belief in their own capabilities to achieve implementation goals. The more confident an individual or group of individuals feel about their ability/abilities, the more likely they are to embrace the intervention and have the commitment to overcome implementation obstacles.
	Individual stage of change: characterisation of the phase an individual is in the change process (e.g. pre-contemplation, contemplation, preparation, action and maintenance).
	Individual identification with organisation: the degree of alignment between the individuals' and the organisation's values and the overall commitment of the individual to the organisation. The higher the level of commitment, the more likely the implementation initiative it is to succeed.
	Other personal attributes: other personal traits such as intellectual ability, motivation, competence and capacity.
Process of implementation	**Planning:** the extent to which the methods and tasks for implementing an intervention are planned. The plan can be formal or informal but should consider all contextual factors.
	Engaging: the extent to which key individuals or groups of individuals (e.g. opinion leaders, formally appointed internal leaders, champions or external agents) are involved. The absence/presence of a leader or their role in the organisation can influence the success of the implementation.
	Executing: the degree to which the different tasks in the implementation plan are completed or done in a timely fashion.
	Reflecting and evaluating: the ability to dedicate time to reflect or debrief before, during or after the implementation as a way to share learnings across the organisation.

Chapter 3. Use of advanced imaging in the management of suspected scaphoid fracture

3.1 Introduction

3.1.1 Clinical condition

The scaphoid bone is an obliquely orientated bone on the radial (thumb) side of the wrist, between the distal carpal row and the radius (Hackney and Dodds 2011). Wrist injury is a common presentation to the Emergency Department (ED) in the UK (Patel et al. 2013). Amongst these patients, the scaphoid is the most commonly fractured carpal bone, accounting for 51-90% of carpal fractures and between 2-7% of all fractures (Yin et al. 2010, Kaewlai et al. 2008; Brooks et al. 2005; Hackney and Dodds 2011). Scaphoid fractures are particularly frequent in young, healthy individuals, usually caused by a fall onto an outstretched hand (Yin et al. 2010, Nguyen et al. 2008).

Reviews have found highly variable estimates for incidence of scaphoid fracture. The systematic review performed by Yin et al. (2010) found a minimum incidence of scaphoid fractures of 5% and a maximum of 50%. However, these values were derived from different reference tests, explaining part of the variation. In addition, inter-hospital inconsistencies in the diagnostic pathway for patients with suspected scaphoid fracture may have further impacted the incidence values retrieved from literature. Taking this uncertainty into account, clinical evidence suggested a confirmed scaphoid fracture incidence value of between 10% and 20% from the overall number of patients presenting with suspected scaphoid fracture (Yin et al. 2010).

3.1.2 The clinical challenge

The management of suspected scaphoid fractures is particularly challenging due to three factors: (i) the low incidence of actual scaphoid fractures in patients presenting with suspected scaphoid fracture; (ii) the limited accuracy of conventional radiograph as the initial imaging modality; and (iii) the potential for complications resulting from a misdiagnosis.

First, of the patients presenting with wrist pain and tenderness at the ED, most do not present with an obvious scaphoid fracture (Patel et al. 2013). It is estimated that the majority, between 66-84%, of patients presenting at the ED will have no definite bone injury at all (Patel et al. 2013; Mallee et al. 2011).

Second, clinical and radiographic diagnosis of scaphoid fracture is often challenging, particularly at the time of presentation. According to Nguyen et al. (2008) this situation could lead to under-diagnosis in up to 40% of cases, and subsequent under-treatment of scaphoid fractures. This would affect patient outcomes, mostly in young people of working age, who most commonly suffer fractures of the scaphoid.

Third, various clinical complications may arise from a misdiagnosed scaphoid fracture. These include non-union, avascular necrosis and secondary wrist arthritis (Bergh et al. 2013). Bearing this in mind, early diagnosis and prompt treatment are essential to improve patient outcomes and reduce the risk of potential complications (Yin et al. 2010). In order to avoid potential clinical complications associated with scaphoid fractures, clinicians tend to over-treat patients with suspected scaphoid fractures, leading to the routine use of splints and even plaster casts in patients with no radiographic evidence of fracture. This can significantly affect patients' quality of life and their overall experience.

In addition to the financial impact from the NHS perspective, scaphoid fracture over-treatment leads to broader societal costs, mainly due to the loss of productivity resulting from the unnecessary use of plaster casts (Yin et al. 2010; Yin, Zhang, and Gong, 2015). Yin, Zhang and Gong (2015) created a decision tree model using published data and concluded that, from a societal perspective, the diagnostic strategy with Computed Tomography (CT) scanning on presentation was the most cost-effective strategy. This was due to the high societal costs, specifically lost productivity, of using immobilisation methods in patients without a scaphoid fracture.

Based on this evidence, the present study evaluated the pathway associated with suspected scaphoid fractures, aimed at providing an early and accurate diagnosis and subsequent appropriate and timely treatment.

3.1.3 Economic evidence: systematic literature review

Despite the known superior accuracy level of advanced imaging in the diagnosis of suspected scaphoid fractures (Yin et al. 2010), it is important to understand its economic impact in the context of real-world clinical practice. To this effect, a systematic review, published by the student, synthesised the economic evidence on the use of advanced imaging, particularly CT and Magnetic Resonance Imaging (MRI), in the management of suspected scaphoid fractures (Rua et al. 2017).

Search strategy

The search strategy was based on the PICOS (Population, Intervention, Comparator, Outcome, Study design) framework, as below.

- **P**opulation/Patient: Patients with suspected scaphoid fracture.
- **I**ntervention/treatment: Advanced imaging (e.g. MRI, CT, Ultrasound or Nuclear Medicine - bone scintigraphy).
- **C**omparator(s): Conventional radiography (x-ray or radiographs).
- **O**utcome(s): primary outcome - total costs associated with the use of advanced imaging in the acute management of suspected scaphoid fractures; secondary outcomes: economic evaluations (e.g. cost-effectiveness analysis).
- **S**tudy Design: Any type of study design.

In summary, this review included studies of any design which evaluated any economic data associated with the use of any advanced imaging modality in the management of patients with suspected scaphoid fractures.

Inclusion criteria

The inclusion criteria for this review were:

- Quantitative study of any design in which the clinical condition was suspected scaphoid fracture.
- Study analysing the use of any kind of advanced imaging modality, particularly CT, MRI, Ultrasound and Nuclear Medicine, with or without a comparison to the traditional use of conventional radiography.
- Study including any form of economic analysis and outcomes, such as healthcare resource use, cost analyses, economic evaluations, lost productivity or quality-adjusted life years (QALYs).

Exclusion criteria

Studies were excluded if the language was not English, Spanish or Portuguese or if the participants were not human. Only studies published from 1990 onwards were included given that formal economic evaluations were rarely conducted prior to this date and, furthermore, the imaging field has substantially changed since then. Studies that only focused on the treatment of confirmed scaphoid fractures were also excluded.

Databases

The following databases were searched: Ovid Classic and EMBASE (1990 to 2016 May 20), Ovid MEDLINE(R) In-Process & Other Non-Indexed Citations, Ovid MEDLINE(R) (1990 to 2016 May 20), Cochrane Library NHS Economic Evaluation Database and Cochrane Library CRD Health Technology Assessment Database (1990 to 2016 May 24). The search strategy, listed in Appendix II, was developed and conducted by the student (TR). The search strategy was consistently used in the databases searched, with only minor adjustments specific to the database searched. The use of truncation, wildcards and Boolean logic, aimed at maximising the number of relevant articles. In addition, references cited in the identified papers were also examined. Duplicates were then removed using the reference manager software Zotero.

Screening and quality assessment

One reviewer (the student, TR) screened titles and abstracts for relevance, whilst a second reviewer (research assistant, SH) assessed a random 20% of exclusions determined by a number generator in Microsoft Excel. Full texts of selected articles were retrieved and independently assessed by two reviewers (TR, SH) based on the defined inclusion and exclusion criteria.

Data were extracted using the data extraction form produced by the National Institute for Health and Care Excellence (NICE). Quality assessment was performed using guidelines on economic quality of economic evaluations issued by NICE (NICE 2012b). Further guidance was obtained from Drummond et al. (2004).

As the studies retrieved presented different methodologies regarding study design, follow-up period, type of imaging modality, type of economic/cost data, economic perspective, it was not deemed appropriate to summarise the evidence using a meta-analysis. Hence, a descriptive synthesis of evidence was undertaken.

Results

The full selection process flow chart for the database searches is depicted in Figure 26. The database searches generated 211 papers: Medline (57); Embase (93); NHS Economic Evaluation Database and Cochrane Library NHS Economic Evaluation Database (61). The screening of references cited in identified papers increased the total list of records to 212. A total of 151 papers (all written in English) remained after removing duplicates. A total of 130 records were excluded following screening of abstracts, leaving a total of 21 full text papers to be reviewed. Of these, 6 records were excluded for various reasons (see Figure 26), leaving a total of 15 relevant papers to be reviewed.

Descriptive analysis

The studies were compared in terms of the: (i) study design; (ii) type of intervention (imaging modality considered); (iii) timing of the intervention (e.g. utilisation of advanced imaging at day 1 or day 7 following the injury); and (iv) type of economic outcome measured. Appendix III summarises the findings from the 15 studies included in this review focusing on the: study design and follow-up duration; intervention(s) and comparator(s); population characteristics and sample size; clinical and economic outcomes; main economic findings; and the author's conclusions.

Country of origin

Over 50% of the research included was conducted in Europe, with four studies from the UK (Patel et al. 2013; Burns et al. 2013; Jenkins et al. 2008; Saxena et al. 2003) and one from the following countries: the Netherlands (Tiel-van Buul et al. 1995); Spain (Moreno Ramos et al. 2013); Norway (Bergh et al. 2013); and Denmark (Hansen et al. 2014). The remaining evidence included studies from Australia (Brooks et al. 2005; Ganeshalingam, Eng, and Page 2013; Kelson, Davidson, and Baker 2016), US (Dorsay, Major, and Helms 2001; Karl, Swart, and Strauch 2015), China (Yin, Zhang, and Gong 2015) and New Zealand (Gooding A., Coates M., and Rothwell A. 2004).

Study design

Three of the studies used randomised designs, with two randomised controlled trials (RCT) (Patel et al. 2013; Brooks et al. 2005) and one randomised controlled pilot (Kelson, Davidson, and Baker 2016). Four quasi-experimental studies, with non-randomised designs, were also identified (Moreno Ramos et al. 2013; Bergh et al. 2013; Hansen et al. 2014; Gooding A., Coates M., and Rothwell A. 2004). Eight studies were based on economic models, with three being economic evaluations (Tiel-van Buul et al. 1995; Karl, Swart, and Strauch 2015; Yin, Zhang, and Gong 2015) and five cost analyses (Dorsay, Major, and Helms 2001; Burns et al. 2013; Jenkins et al. 2008; Saxena et al. 2003; Ganeshalingam, Eng, and Page 2013).

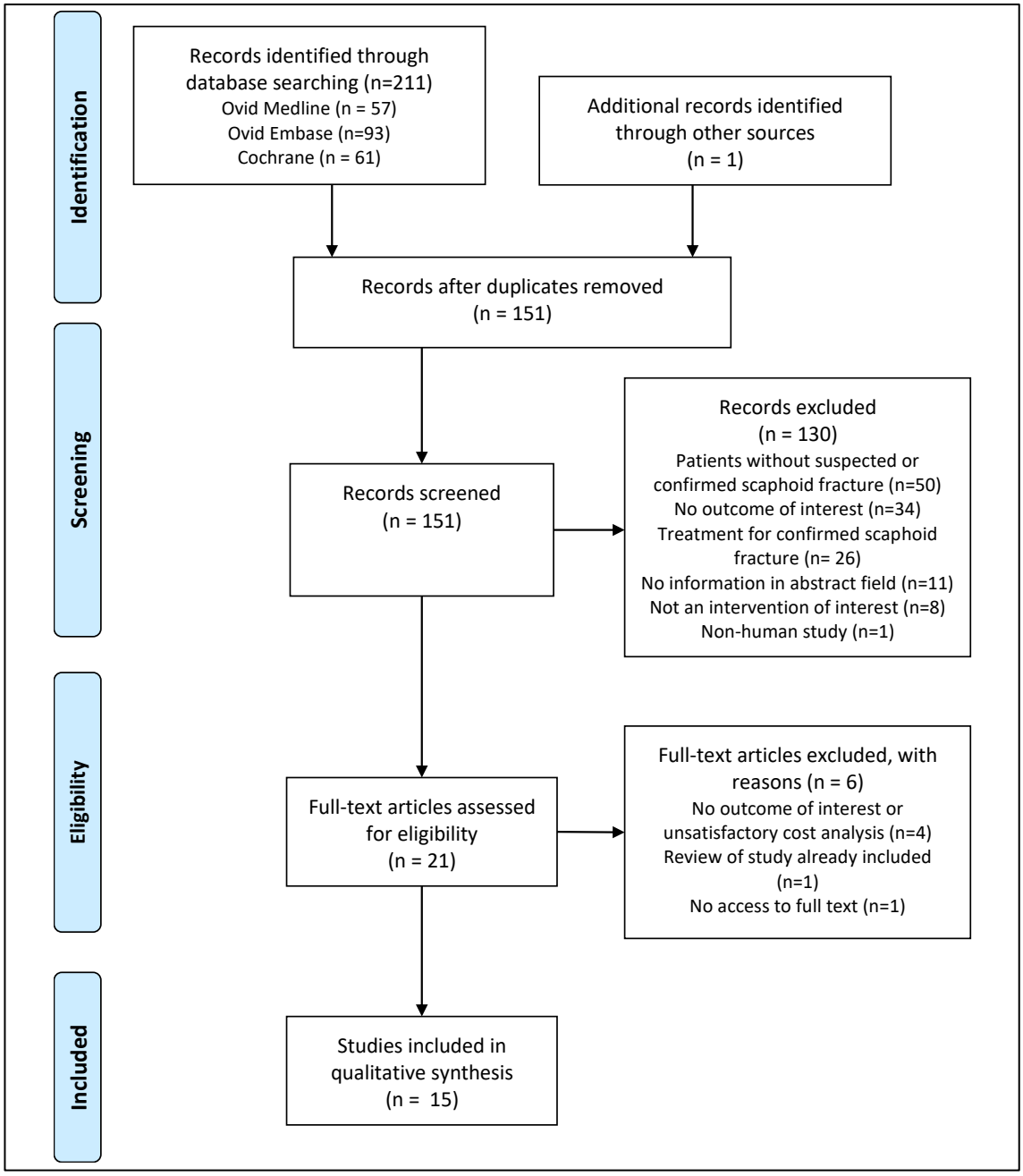


Figure 26. PRISMA flow chart summarising the selection process of relevant studies.

Study samples

The studies appeared to be based on comparable populations: patients with suspected scaphoid fractures but without findings suggestive of scaphoid fracture on the initial radiographs (x-ray). Sample sizes for the three randomised studies ranged from 16 participants (Kelson, Davidson, and Baker 2016) to a total of 84 (Patel et al. 2013).

Imaging Modality

Out of the 15 studies reviewed, all but Tiel-van Buul et al. (1995) studied the inclusion of MRI in the management of suspected fractures. However, some studies also included other imaging modalities, specifically Bone Scintigraphy (4 studies), CT (3 studies) and Ultrasound (1 study).

Time of Intervention

Along with the type of imaging modality used, the timing of the diagnostic test was of vital importance as the same imaging modality used at different points in time might produce different clinical and economic outcomes. The three randomised clinical studies considered the use of advanced imaging between 2-5 days (Brooks et al. 2005), 1-3 days (Kelson, Davidson, and Baker 2016) and within 2 days (Patel et al. 2013) following the injury. Other studies, particularly economic modelling studies, broadly evaluated the use of advanced imaging on the day of injury and up to 2 weeks following the injury. As an example, Yin et al. (2015) found that the immediate use of CT was the most cost-effective strategy for managing suspected scaphoid fractures (Yin, Zhang, and Gong 2015).

Economic evidence

The analysis of economic outcomes were grouped according to the respective study design and perspective of analysis (either healthcare payer or societal perspectives). Furthermore, the distinction between cost analyses and cost-effectiveness and cost-utility analyses was also noted.

The three studies that used **randomised** allocation methods did not find statistically significant differences in costs from a healthcare payer's perspective. Two randomised controlled trials, by Patel et al. (2013) and Brooks et al. (2005), and one randomised controlled pilot, by Kelson et al. (2016), found no statistically significant differences in terms of total mean or median costs due to the use of MRI in the management of suspected scaphoid fractures [£504.13 intervention group vs £532.87 control, p=0.9 (Patel et al. 2013); median cost of $411.48 intervention group vs $296.42 control group, p=0.19 (Brooks et al. 2005); and median cost of $335.81 intervention group vs $337.09 control group, p=0.74 (Kelson, Davidson, and Baker 2016)]. This seemed to be due to the fact that all three studies appeared to be either underpowered, with small sample sizes, or not powered to detect statistical significant differences in economic outcomes [e.g. one RCT was powered to detect differences in days unnecessarily immobilised (Brooks et al. 2005)]. However, when potential societal costs were included, in particular costs associated with time off work, the use of advanced imaging, in this case MRI, was likely to be cost saving (Kelson, Davidson, and Baker 2016) or cost-effective (Brooks et al. 2005). In one of the RCTs, by Brooks et al. (2005), a cost-effectiveness analysis estimated that the use of MRI saved $30.8 (95% CI $2.97 to $69.94) per day due to the prevention of unnecessary immobilisation (Brooks et al. 2005).

Evidence from the four **non-randomised** empirical studies was, from a healthcare payer's perspective, variable but exhibited a trend of increased healthcare costs associated with the use of advanced imaging. Three studies found that standard treatment (i.e. with wrist immobilisation and no use of early advanced imaging) was less expensive than the strategy where early advanced imaging was available (Moreno Ramos et al. 2013; Hansen et al. 2014; Gooding A., Coates M., and Rothwell A. 2004). It is however relevant to note that two studies (Moreno Ramos et al. 2013; Gooding A., Coates M., and Rothwell A. 2004) did not perform a statistical analysis (no p-value or confidence intervals) around the economic outcomes. Bergh et al. (2014) found no statistically

significant cost-differences in the management of scaphoid fractures. Out of the four quasi-experimental studies, only Gooding and colleagues (2004) performed an economic evaluation (the remaining three focused on cost analyses). Findings from this study suggested that advanced imaging, in this specific case MRI, despite having increased the total hospital costs, was cost-effective with an average cost of $692 to exclude a scaphoid fracture (Gooding A., Coates M., and Rothwell A. 2004). Finally, as with the RCT studies, the two studies that assumed a societal perspective reported that the utilisation of advanced imaging significantly reduced societal costs (Bergh et al. 2013; Hansen et al. 2014). Bergh et al. (2014) estimated that indirect costs for employees (due to sick leave) accounted for up to 85% of the total management costs.

Three studies used **economic modelling** to perform **cost-effectiveness** analyses around the use of advanced imaging in the management of suspected scaphoid fractures (Tiel-van Buul et al. 1995; Karl, Swart, and Strauch 2015; Yin, Zhang, and Gong 2015). Yin, Zhang, and Gong (2015) evaluated six potential interventions using a decision tree model. These interventions varied in the imaging modality and also in the timing of when the actually imaging scan was performed. Cost-effectiveness analyses showed that immediate CT and day-3 MRI were the most cost-effective strategies (Yin, Zhang, and Gong 2015). A second study, by Karl et al. (2015), using QALYs as the measure of effect and taking a societal perspective, found that advanced imaging was dominant over empiric cast immobilisation. This study also found the use of MRI over CT presented an incremental cost-effectiveness ratio (ICER) of $41,000/QALY (£27,350/QALY). Depending on the healthcare payer's willingness to pay thresholds and also local institutional costs and imaging availability, a strategy to manage suspected scaphoid fractures using MRI was likely to be both feasible and cost-effective (Karl, Swart, and Strauch 2015). A third study, by Tiel-van Buul et al. (1995), evaluated the use of bone scintigraphy. The incremental costs incurred to save one non-union by using bone scintigraphy were found to be one-third of the price of repeated radiography at 6 weeks. The authors concluded that the intervention was cost-effective, although no specific ICER was presented (Tiel-van Buul et al. 1995).

Five studies used **economic modelling** to perform **cost analyses** (Dorsay, Major, and Helms 2001; Burns et al. 2013; Jenkins et al. 2008; Saxena et al. 2003; Ganeshalingam, Eng, and Page 2013). These studies reported divergent findings concerning the cost impact of using advanced imaging in the management of suspected scaphoid fractures. One study presented favourable cost evidence (Saxena et al. 2003), whilst another one found no cost difference (Dorsay, Major, and Helms 2001) and four studies reported unfavourable evidence (Burns et al. 2013; Jenkins et al. 2008; Saxena et al. 2003; Ganeshalingam, Eng, and Page 2013). Saxena et al. (2003) hypothesised five potential interventions using advanced imaging (MRI or bone scintigraphy) at different timings (on day 1, within a few days or in 2 weeks). Depending on the type of imaging modality, the use of advanced imaging led to an increase or decrease in total management costs (Saxena et al. 2003). Two interventions were found to be cost saving, particularly MRI on day 1 and, to a lesser degree, MRI within few days followed by a review of results on the same day (Saxena et al. 2003). One study, by Dorsay et al. (2001), found no significant cost difference

between the strategy with screening MRI and the comparator with conventional immobilisation. In contrast, four studies found that the use of advanced imaging led to increased healthcare costs. Saxena et al. (2003) found that the use of advanced imaging was more expensive if the review of the imaging results was not performed on the same day of the scan as it led to an increased number of outpatient appointments. Burns et al. (2013) concluded that the costs of management of suspected scaphoid injuries was higher than the additional cost of performing an MRI scan by $139 (£97) (Burns et al. 2013). Jenkins et al. (2008) established that the use of advanced imaging, particularly MRI and CT, led to an increased costs of $225 (£158) and $82 (£58), respectively (Jenkins et al. 2008). Finally, Ganeshalingam et al. (2013) reported that the use of MRI led to an increase of $242 (£168) in costs associated with the management of suspected scaphoid fractures. However, as with other studies that included societal costs due to time off work, Ganeshalingam et al. (2013) found that the use of MRI led to savings of $1,655 (£1,151) per patient.

Summary of the systematic review:

The systematic review published by the student (Rua et al. 2017) identified a number of different methodologies such as cost, cost-effectiveness and cost-utility analyses. The findings from different studies might not be directly comparable as the country of origin, study design, type of imaging, timing after injury and also the economic outcomes varied considerably between studies. In addition, papers reviewed included important limitations such as inappropriate reporting methods, short follow-up periods and statistical under-powering. Moreover, all three economic modelling studies that evaluated the use of immediate advanced imaging (i.e. on the day of injury) reported divergent findings and shared important methodological issues, particularly the lack of empirical evidence needed to estimate health care resource use following the use of immediate advanced imaging. The evidence presented should therefore be interpreted with caution. Taking this into consideration, the systematic review found economic findings differences based on the perspective of analysis (Rua et al. 2017). If a societal perspective was taken, the use of advanced imaging was, regardless of the timing and type of intervention, likely to generate cost savings. In contrast, if a healthcare perspective was assumed, the studies reviewed in the systematic review did not provide conclusive economic evidence (Rua et al. 2017), although the randomised studies did not find statistically significant cost differences. The non-randomised studies, however presented two major limitations: (i) lack of empirical data to measure the impact of advanced imaging on healthcare resource use (e.g. no resource use was considered following a negative finding on the advanced imaging scan); and (ii) the economic models considered did not necessarily reflect real-world clinical practice (e.g. the use of immediate advanced imaging did not lead to the detection of other bone fractures or soft-tissue injuries).

Based on this evidence, we concluded that there was a need for a well-designed economic study to assess the clinical and economic impact of using immediate advanced imaging in the management of suspected scaphoid fractures based on empirical data.

3.1.4 Standard care

The clinical and radiographic diagnosis of scaphoid fractures is often challenging at the time of presentation (Yin et al. 2010; Mallee et al. 2011). This is particularly true if imaging protocols are limited to conventional radiographs and exacerbated by the lack of standardisation of imaging protocols for suspected scaphoid fracture across hospitals (Yin et al. 2010; Mallee et al. 2011).

Consistent with other healthcare providers, the current diagnostic pathway at GSTT includes an initial clinical assessment on arrival (usually via the ED). Subsequently, if a scaphoid fracture is suspected, plain radiographs are performed, using a specific, 4-view, scaphoid protocol.

Based on the results, the clinical pathway and subsequent treatment varies:

- **Positive** findings (i.e. abnormal findings on the initial 4-view plain x-ray):

Patients are given a splint and are referred to an initial fracture clinic consultation between 1-2 weeks after the initial ED attendance. Subsequently if a scaphoid fracture is confirmed, the patient's arm is put in a plaster cast.

At the time of their fracture clinic appointment, the majority of patients with a confirmed scaphoid fracture undergo a CT scan (or to a lesser degree, an MRI scan) to establish whether the fracture is displaced and, if so, the degree of displacement. Based on these imaging findings, the patient's scaphoid fracture is managed in one of two ways: (i) plaster cast for a period of at least 6 to 8 weeks; or (ii) to a lesser degree, surgery in the case of displaced fractures or proximal pole fractures.

- **Negative** findings (i.e. no abnormal findings on the initial 4-view plain x-ray):

Similar to patients with positive findings, patients with negative findings on the initial radiographs are referred to a fracture clinic consultation 1-2 weeks after the initial ED attendance. At this point repeated conventional radiograph (usually 4-view plain x-ray) and, to a lesser degree, a CT scan (or an MRI scan) is carried out. Based on these imaging findings, the patient is either discharged or is given the treatment as above (i.e. either plaster cast or surgery). At subsequent follow-up appointments, if the 4-view plain x-ray is negative but the patient remains symptomatic, further imaging (usually a CT scan) is carried out (if not performed on the initial outpatient appointment).

As summarised above, the management of people with scaphoid fractures at GSTT comprised the use of several imaging technologies including conventional radiograph and, in some cases, advanced imaging (CT and, to a lesser degree, MRI). The rationale for the use of CT, rather than MRI, seems to be explained by the limited availability of MRI. In fact, CT scans can generally be carried out and reviewed by the referring clinician on the day the patient is seen in the fracture clinic. However, as with any radiation based imaging technique, the utilisation of CT is not harmless as exposure to radiation which is associated with both non-stochastic and stochastic detrimental health effects.

3.1.5 Proposed Intervention

The proposed intervention was based on the use of immediate MRI, i.e. during the acute episode in the ED, as a decision tool for the management of suspected scaphoid fractures. Despite the higher cost and reduced availability of MRI when compared to CT, this decision was based on both the lack of radiation of MRI and the superior accuracy levels of MRI in the management of suspected scaphoid fractures (Rua et al. 2018).

The results from the MRI subsequently informed the diagnostic and treatment pathway as follows:

- **Positive** findings (i.e. abnormal findings on the initial MRI):

Positive findings such as scaphoid or any other bone fractures were treated with plaster cast during the ED episode and referred to the next available fracture clinic appointment where clinical and radiographic follow-up is considered. Given the MRI's very high sensitivity to rule-in scaphoid fractures among suspected scaphoid fractures, it was hypothesised that the use of MRI would lead to an accurate detection of scaphoid (or other bone) fractures on presentation to the ED.

- **Negative** findings (i.e. normal findings on the initial MRI):

Patients with negative findings for bone fracture (scaphoid or any other bone) or no major soft tissue injury were discharged from the ED with no formal follow-up at Fracture Clinic. However, given that participants without fractures may still have wrist pain 2 weeks post-injury, to further diagnosis or treatment of symptomatic patients should be considered (see section 3.2.5 for further detail). Given the very high specificity of MRI to rule-out scaphoid fractures among suspected scaphoid fractures, it was hypothesised that the use of MRI would lead to the safe discharge of patients without any scaphoid or any other bone fracture.

3.1.6 Rationale for the trial

Economic evidence around the use of advanced imaging (particularly MRI) in the assessment of suspected scaphoid fractures comes from economic modelling studies, quasi-experimental studies and three randomised controlled studies. However, in all empirical studies the use of MRI was not performed as a first line investigation during the initial acute episode. This trial considered the use of immediate MRI, i.e. during the initial presentation to the ED. To the best knowledge, this approach is innovative and could provide a foundation on which to base national and international best practice. The rationale and design for this trial was published in the Journal of Clinical Trials, whilst its results were published in the Bone & Joint Journal and Value in Health journal (Rua et al. 2018).

3.2 Methods

3.2.1 Aims, Objectives and Hypotheses

Aims of the study:

The aim of the SMaRT (**S**caphoid **Ma**gnetic **R**esonance Imaging in **T**rauma) trial was to assess clinical effectiveness and conduct cost and cost-effectiveness analyses around the use of immediate MRI in the management of suspected scaphoid fractures compared to conventional management with immobilisation and clinical and radiographic follow-up.

Study objectives:

One primary and seven secondary objectives were considered.

Primary Objective: To estimate the 3-month costs associated with two clinical pathways in the ED: (a) the control group, with conventional radiograph as the only imaging modality in the ED; or (b) the intervention, a hybrid approach that considers the use of wrist MRI in the ED as an add-on test for patients with negative findings on the initial conventional radiograph.

Secondary Objectives:

I. To perform a cost analysis at 6 months associated with the intervention group compared to the control group.

II. To perform cost-effectiveness analyses at 3 and 6-months to estimate the incremental cost-effectiveness ratio (ICER) associated with the proposed intervention compared to the control group. Two measures of effect were considered: (a) pain levels; and (b) QALY.

III. To estimate the mean cost per correctly diagnosed scaphoid fracture in the intervention group compared to the control group.

IV. To assess the overall patient satisfaction associated with the intervention group compared to the control group.

V. To estimate the accuracy associated with the proposed intervention (i.e. immediate wrist MRI) in the detection of scaphoid fracture compared to the current pathway (i.e. 4-view conventional radiographs).

VI. To estimate the time taken to reach a definitive diagnosis in the intervention group compared to the control group.

VII. To estimate the amount of time off work or informal care needs due to the suspected scaphoid fracture in the intervention group compared to the control group.

Hypotheses:

Primary Hypothesis:

The primary hypothesis for the study was that the addition of immediate MRI in the management of suspected scaphoid fractures with negative findings on the initial conventional radiograph will decrease the overall 3-month NHS costs per patient compared to those in the standard care.

The standard care pathway, i.e. the control group, includes the use of the initial x-ray only as part of the emergency diagnostic pathway, as opposed to the intervention group where an add-on wrist MRI test was performed. The 3-month timeline was considered appropriate as all relevant costs and benefits should be realised within this timeframe.

The underlying rationale was that the early use of a more expensive diagnostic tool (i.e. MRI) will enable an improved accuracy in the diagnosis of scaphoid fracture and change the subsequent treatment. This may avoid unnecessary treatment as up to four out of five patients without a fracture are immobilised for long periods of time, and ultimately avoid downstream costs associated with fracture clinic appointments and repeated diagnostic tests.

Secondary Hypotheses:

Additionally, it was hypothesised that the intervention will:

- Reduce total 6-month NHS costs.
- Be cost-effective at 3 and 6-months, using pain levels and QALYs as the measure of effect.
- Be cost-effective at 3 months, using the number of correctly diagnosed scaphoid fractures as the measure of effect.
- Increase levels of patient satisfaction.
- Improve diagnostic accuracy in the detection of scaphoid fractures.
- Reduce the time taken to reach a definitive diagnosis (i.e. to either rule in or rule out a scaphoid fracture).
- Reduce the amount of time off work, informal care or hand immobilisation with plaster cast due to the suspected scaphoid fracture.

3.2.2 Study design

Randomised Pragmatic Trial

The SMaRT trial was a single-site, pragmatic, prospective, parallel, non-blinded, randomised trial. The SMaRT trial was designed as a pragmatic trial, aimed at assessing the real-world effectiveness of the intervention applied as part of routine clinical practice to a heterogeneous population.

The SMaRT trial presented two key innovative features. First, the use of MRI as part of the initial visit to the acute setting (in the ED) was, to our best knowledge, novel. This approach, although operationally challenging, aimed simultaneously at: (i) reducing the number of healthcare contacts as patients with negative findings on the MRI might not need further appointments; (ii) improving

overall patient satisfaction by using a more definitive approach; and (iii) increasing overall NHS efficiency by promoting decisive care and reducing the need for unnecessary care (e.g. follow-up of patients with no scaphoid fracture). Second, the intervention proposed a holistic transformation of the diagnostic and treatment pathways as a direct consequence of the MRI findings. As an example, only a proportion of participants with negative findings on the MRI (i.e. no scaphoid or other bone fractures) required any form of follow-up at secondary care. Indeed, participants with no relevant MRI findings that remain pain-free after two weeks were not expected to require any follow-up at all (apart from the 3 month 4-view plain x-ray for the purposes of research). Hence, by using a more advanced and accurate imaging modality earlier in the clinical pathway we were able to appropriately change the participant's subsequent care.

The SMaRT trial was designed to detect differences in resource use between the intervention (immediate wrist MRI) and hospital standard care (control group) on an intention to treat basis. The trial's design explicitly aimed to minimise potential sources of biases and maximise the generalisability of our results. First, to reduce the selection and/or population bias, block randomisation was performed by an external organisation (King's Clinical Trial Unit). The use of random and concealed allocation methods is the most rigorous method to test the direct effect of the intervention on outcomes (Dettori, 2010). Second, findings from the SMaRT trial may be transferrable to other UK-based hospitals due to two factors: (a) the control group pathway largely reflects standard care in the UK; and (b) the trial's inclusion criteria is broad, reflective of the overall population undergoing evaluation of suspected scaphoid fractures.

Participant Groups

Participants were randomly allocated to two groups: the control group (standard care); and the intervention group (MRI group). Participants in the control group followed standard care that relied in the use of conventional radiograph as the only diagnostic tool as part of the ED episode. Participants randomised to the intervention group underwent a wrist MRI scan as part of the ED episode as an add-on to the initial conventional radiograph. This MRI scan was performed based on abbreviated imaging protocols compared to a conventional non-acute wrist MRI scan. The rationale for this approach derived from the combination of two reasons: first, not to overburden the acute diagnostic pathway; and second, to obtain enough clinical information to accurately rule-in or rule-out a major wrist injury (e.g. scaphoid fracture, other bone fracture and soft-tissue injuries).

Random Allocation

Randomisation method:

Participants were assigned to groups using a web-based automatic 1:1 block randomisation sequence generated by an external organisation, King's Clinical Trial. A fixed size block of 10 patients, except for the last block (size of 6 due to the sample size), was used.

Allocation Concealment:

The randomisation method was based on a web-based system hosted by an organisation exterior to the SMaRT trial. Hence, the allocation sequence was fully concealed from both the participants and any member of the research staff.

Blinding:

The SMaRT trial presented a non-blinded design. Given the nature of the intervention and its impact on subsequent care it was not deemed feasible to blind participants or research staff to the intervention.

This lack of blinding might lead to potential conscious or unconscious performance bias. However necessary, this constituted a SMaRT trial limitation. As the latter had the potential to affect the primary outcome (e.g. over or underutilisation of healthcare resources), strict clinical pathways were disseminated prior to the trial's initiation.

In addition, the lack of blinding might have led to attrition bias, i.e. different attrition rates associated with both randomisation groups. Different preventive steps were put in place to mitigate this risk, particularly: (a) ensuring good communication between participants and different members of the research team; (b) financial incentives for participants to complete the study follow-up; (c) utilisation of databases that were not based on participants' self-reporting data; and (d) use of intention-to-treat analysis (i.e. all participants were analysed as per their respective randomisation group regardless of any other event) (Tal 2011; Peacock, Kerry, and Balise 2017).

Follow-up Period

All participants were followed up to a period of 6 months. Data were collected at baseline and at 1, 3 and 6 months post-recruitment.

3.2.3 Ethical Approval, Trial Registration and Funding

The Health Research Authority and Research Ethics Committee (South East Coast – Surrey REC) approved the SMaRT trial research on the 17th May 2016. The REC reference was 16/LO/0826 and the IRAS project ID was 180601. The SMaRT trial commenced on the 15th June 2016, with the first participant being recruited on the 06th July 2016. The SMaRT trial was registered on clinicaltrials.gov (Clinical Trial Registration: NCT02801149) on the 15th June 2016.

The SMaRT trial was fully funded by a grant secured from Guy's and St Thomas' Charity.

3.2.4 Selection, withdrawal of participants and sample size

Study Setting:

The study was conducted at an independent single Trust site, at GSTT. Participants were recruited from the ED.

Inclusion Criteria:

Patients were considered to be eligible for the study if at least one of the inclusion criteria was met and none of the exclusion criteria was present. Any patient aged 16 years or over presenting at the ED with clinical history and examination consistent with suspected scaphoid fracture with negative findings on the initial 4-view plain conventional radiograph were considered eligible (Figure 27). It was considered that a patient presented a suspected scaphoid fracture if at least one of the following criteria was present:

- Isolated pain / tenderness over the Anatomical Snuff Box or Scaphoid Tubercle or pain in the scaphoid region during axial loading of the 1st metacarpal.
- History of recent fall (< 14 days) on an outstretched hand, wrist injury or poor history associated with examination findings suggestive of scaphoid fracture.

Figure 27. 4-view scaphoid initial radiographs at the ED showing no evidence of a scaphoid or any other bone fracture. Hence, the participant was deemed eligible for the SMaRT trial.

Exclusion Criteria:

Patients were considered to be ineligible if at least one of the exclusion criteria was present:

- Patients presenting outside GSTT catchment area who were not willing to be followed-up at GSTT;
- Patients with a confirmed scaphoid fracture following the initial radiograph exam;
- Patients with confirmed ipsilateral upper limb injury/injuries (e.g. wrist/forearm/arm injury) following the initial conventional x-ray examinations, regardless of the findings around the suspected scaphoid fracture.
- Patients with suspected scaphoid fracture not admitted through the ED;
- Patients who lacked capacity to give consent or participate in the study;
- Patients that were already taking part in a Clinical Trial of an Investigational Medicinal Product;
- Pregnancy;
- Patients screened for the study at the ED on weekdays before 7.30 am and after 6 pm;

- Patients screened for the study at the ED on weekends or Bank Holidays before 9 am and after 4 pm.

Contrary to previous studies, patients with previous scaphoid or wrist injuries were included in the trial. The last two exclusion criteria related to operational challenges associated with the limited time provision of MRI services.

Criteria for Premature Withdrawal:

At all times, particularly when any follow-up was due, the willingness of participants to take part in the study was reassessed. If participants gave a reason for their withdrawal, this was recorded, however participants did not need to provide a reason for withdrawing the study.

Losses to Follow-Up:

If a patient moved to outside GSTT's catchment area, the research staff made every effort to ensure that the participant was still followed-up. In any case, and in order to decrease the probability of losses to follow-up, all participant's GP were, as per the informed consent, contacted to obtain data regarding their primary care and secondary care (whenever outside GSTT's remit) resource utilisation. Only if the participant expressed their wish to withdraw the study, he/she was withdrawn from the study. Given the high mobility within the Greater London area, it was estimated that up to 50% of the participants enrolled in the study could be lost to follow-up.

Sample Size:

For the purpose of the primary objective, the estimated sample size was calculated based on several parameters or assumptions:

- Test family: 2-sided t test
- Statistical test: means – difference between two independent means (two groups)
- Type of power analysis: *a priori*
- Average cost per patient in the current pathway: £325 (estimated using GSTT unit cost data);
- Average cost per patient in the proposed pathway: £225 (estimated using GSTT unit cost data);
- α err probability: 0.05;
- Power (1- β err probability): 0.85;
- Allocation ratio N2/N1: 1/1;
- Standard deviation of average cost per patient in the current pathway: £150 (assumed to represent about half the average cost per patient);
- Standard deviation of average cost per patient in the proposed pathway: £115 (assumed to represent about half the average cost per patient).

Based on the above parameters, the required sample size was 68 patients. Subsequently, it was considered that 50% of the patients enrolled could be lost to follow-up. Hence, in order to guarantee the 68 patients, the study planned to recruit a total of 136 patients (68 in each group). No interim analyses were planned or performed during the trial.

3.2.5 Interventions

All participants with suspected scaphoid fracture underwent an initial conventional radiograph (scaphoid views) as part of their standard care. If positive findings for fracture were found, the participant was deemed ineligible for the SMaRT trial. If no fracture of the scaphoid or any other bone was seen in the initial x-ray, the patient was deemed eligible to take part in the SMaRT trial.

The rationale behind this design was based on the conventional radiograph's high sensitivity for the detection of scaphoid and other bone fractures, i.e. ability to rule-in, not rule-out, scaphoid fractures. Hence, the authors decided that immediate MRI should be used as an add-on test following negative initial conventional radiographs. The subsections below detailed the diagnostic and, if needed, treatment pathway for both groups of the trial.

Control Group (standard care):

Participants randomised to the control group followed current standard care, i.e. no further imaging tests in the ED (see Figure 28a). Following the initial x-ray, participants were put in a wrist splint prior to the discharge from the ED and booked into a Fracture Clinic appointment in 1-2 weeks, where further imaging scans (usually CT and/or, to a lesser degree, MRI) might be performed. Formal follow-up was deemed appropriate due to the limited ability of conventional radiograph to effectively rule out a scaphoid fracture on presentation to the ED.

In the **absence** of fractures (scaphoid or otherwise) visible on any imaging test, participants were usually discharged following the initial Fracture Clinic. The latter might not be the case if, although no fracture was visible on any follow-up imaging test, participants were still symptomatic. In this case, participants were periodically followed-up in subsequent outpatient appointments where further imaging tests might be performed.

In the **presence** of fractures, scaphoid or otherwise, visible on any imaging test, participants were immobilised using a below elbow plaster cast. Then participants were followed-up periodically up to a period of 4-8 weeks, with the combined use of scaphoid-view x-rays and clinical follow-up with a specialist. This approach was designed to ensure that any fracture had healed prior to discharge.

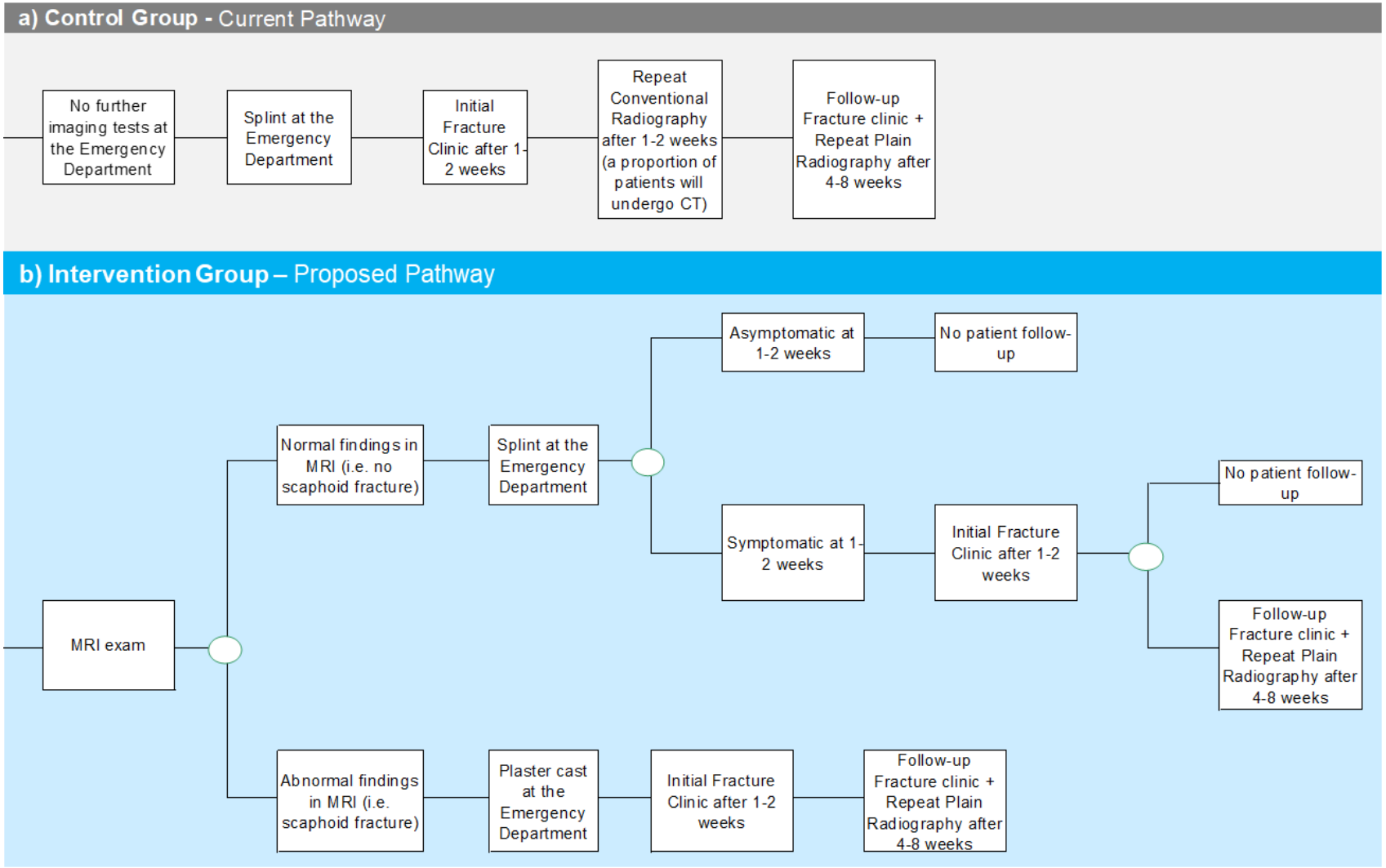


Figure 28. Diagnostic and, if needed, treatment pathway for participants randomised to: (a) the control group; and (b) the intervention group.

Intervention Group (MRI group):

Participants randomised to the intervention group did not receive standard care as they underwent a wrist MRI scan following the initial negative conventional radiograph. An abbreviated wrist MRI, conducted during the ED episode, included two coronal and one sagittal plane (Figure 29). The MRI images were interpreted and reported by the radiology team on a 'live' basis. The MRI images and respective report were released and made available to the ED referrer, who then defined the patient's care as follows and depicted in Figure 28b.

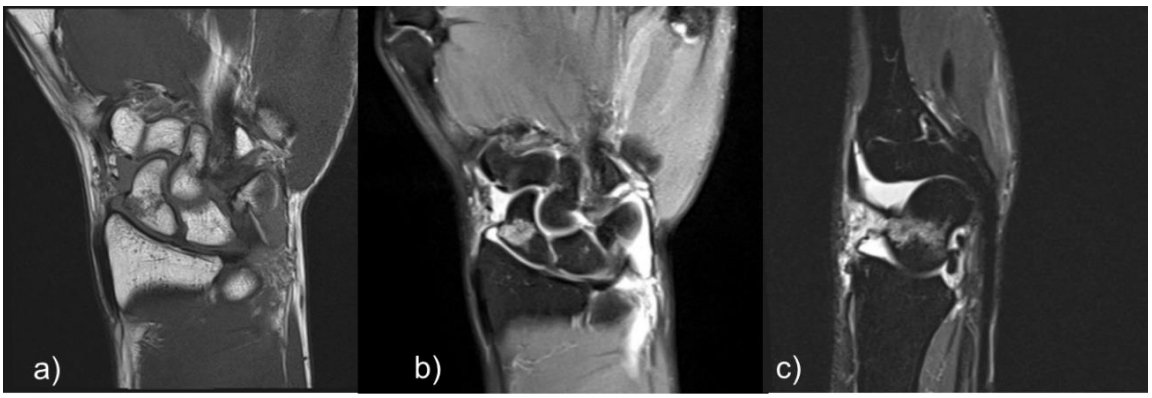


Figure 29. Sequences used in the short-sequence wrist MRI (two coronal and one sagittal plane).

Participants with **no fracture** or major soft tissue injuries (e.g. ligament rupture) visible on the MRI scan were discharged from the ED with a wrist splint and given a contact card. No formal outpatient appointment was booked unless the participant remained symptomatic after 1-2 weeks and, as a consequence, contacted the hospital. In this case, participants were asked to come to a wrist pain clinic appointment.

Participants with a **fracture** or major soft tissue injury visible in the MRI scan had their arm immobilised in a plaster cast prior to the ED discharge and a Fracture Clinic appointment was booked. As with participants in the control group, the majority of confirmed scaphoid or other bone fractures required follow-up appointments and 4-8 weeks immobilisation with plaster cast with a small proportion of displaced fractures requiring surgery.

The MRI scan was interpreted and reported promptly in order to accommodate the current ED targets (4-hour from admission to discharge from the ED), and so as not to delay the diagnosis and, if needed, treatment. In the event that the MRI report was not available for operational reasons during the ED episode, participants were treated and discharged with a supportive splint as per the initial conventional radiograph findings. This approach was consistent with current clinical practice so participants randomised to the intervention group (i.e. with MRI) were not disadvantaged and did not receive inferior care by taking part in this trial. Once the MRI report was available, participants were contacted to receive the results and, when applicable, to schedule subsequent appointments or treatment.

3.2.6 Study Procedures

Trial promotion:

The implementation of a randomised trial in an acute setting was expected to be challenging, particularly due to the short timeline associated with any research-related task during the acute episode and the real-world complications arising from the implementation of a novel diagnostic and treatment pathway that involves different stakeholders (Emergency, Radiology and Orthopaedic Departments). Taking this into consideration, a few weeks prior to the trial start date, the student and other research staff delivered an extensive training programme to over 50 Emergency Nurse Practitioners (ENPs). This training programme provided an overview of the trial with a particular focus on the trial's rationale, design and inclusion and exclusion criteria. At the end of the training programme, the ENPs were given ample opportunities to ask questions and, if happy to take part in the trial, were asked to sign the trial's delegation log. In the context of a trial conducted in an increasingly strained ED, it was considered essential to have the buy-in of key routine care stakeholders for the timely identification of potential eligible participants and, if possible, complete the consent and recruitment processes. Regular updates and one-to-one or group meetings were also implemented on an ongoing basis, particularly when there was an unexpected decrease in the proportion of potential eligible participants being screened.

Screening and Recruitment Procedures:

All participants with suspected scaphoid fracture underwent wrist radiographs (scaphoid views) as part of their standard care. If positive findings for fracture were found, the patient was deemed ineligible for the SMaRT trial. If no fracture were seen on the initial radiographs, the patient was deemed as eligible. This assessment was performed by a clinician from the routine care team, usually an ENP, as part of the ED triage process. Subsequently, the trial was discussed with the patient. These tasks were performed by the ENP who screened the patient, with or without support from the research staff (with or without the student). During these screening procedures, the recruiter was responsible for assessing the participant's capacity to provide informed consent. This process ensured that all participants: (a) were given the updated approved version of the patient information sheet; (b) were fully informed about the study, including its risks and benefits; and (c) confirmed their willingness to participate by signing the approved informed consent. For patients randomised to the intervention group, i.e. with MRI, all GSTT safety and MRI consent procedures associated were subsequently completed.

The flowcharts below describe the processes associated with the recruitment of participants to the control (Figure 30) and intervention groups (Figure 31).

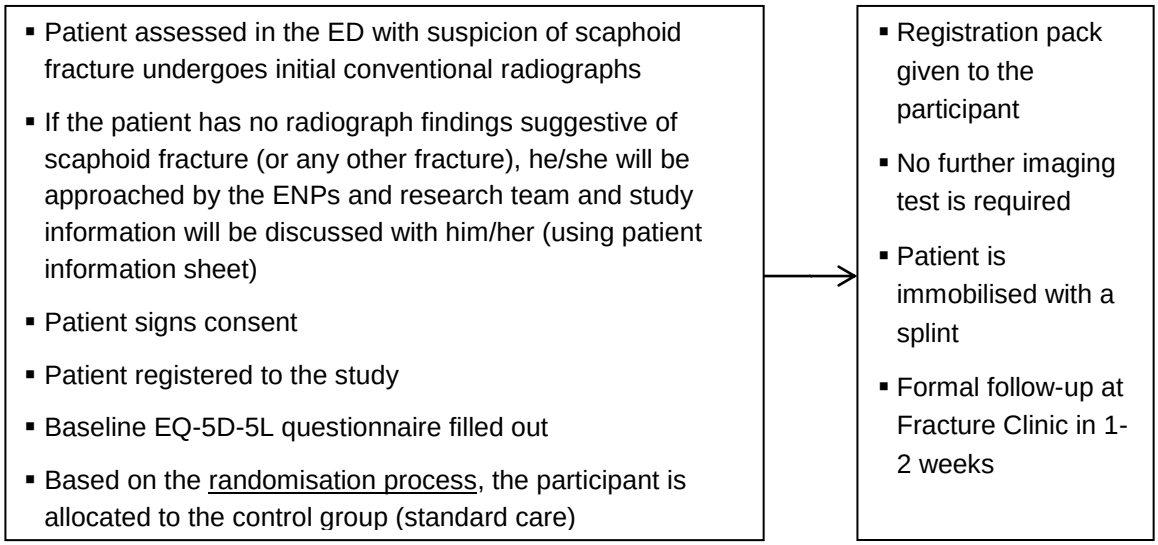


Figure 30: Processes conducted in the recruitment of participants to the control group.

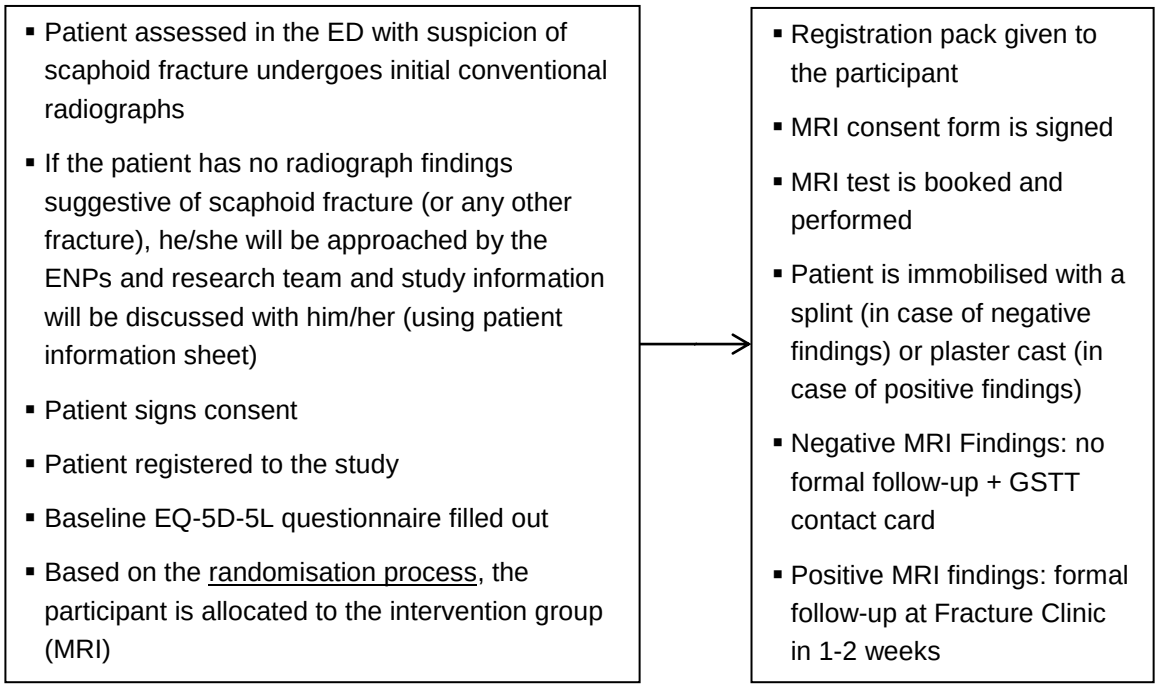


Figure 31: Processes conducted in the recruitment of participants to the intervention group.

In order to avoid potential biases, each participant was asked to complete a baseline EQ-5D-5L questionnaire prior to any randomisation process. This validated questionnaire assessed the participant's self-reported quality of life and was used for the purposes of the economic evaluation.

The diagnostic pathway was not delayed for more than the minimum amount of time necessary to provide the patient with the study information (estimated to be less than 10-20 minutes). Given the patient's non-critical condition and the timely recruitment process, it was not anticipated that this task would have any negative impact on patient outcomes or any performance outcome associated with GSTT. Furthermore, it was explained to the participant that they were free to withdraw from the study at any point in time.

Any participant withdrawals were recorded in the screening log alongside with the reason of withdrawal, if given.

Follow up Procedures

The follow-up procedures are presented in Table 8 and Figure 32).

Participants enrolled in the study were invited to undergo 4-view wrist radiographs and a research consultation with an Orthopaedic Consultant 3 months (±2 weeks) after the initial ED presentation. This exam and specialist appointment confirmed whether a: (a) scaphoid or any other bone fracture healed appropriately; or (b) scaphoid or any other bone fracture was missed. The 3-month radiograph was used as a reference to estimate the accuracy levels in the intervention and control groups.

Participants were asked to complete a validated questionnaire (EQ-5D-5L) at three points in time: month 1, 3 and 6 post-recruitment. A second non-validated questionnaire was developed to assess patient experience in both the intervention and control group at 3 months post-recruitment. Once completed, all questionnaires were posted to GSTT using a pre-paid envelope or, if preferred by the patient, emailed to GSTT. Furthermore, all participants were provided with a diary in the registration pack. In these diaries, participants were asked to record any scaphoid-related hospital and GP visits, community care, medications and investigations. This diary was filled out weekly for a period of 13 weeks (i.e. 3 months) and the participant handed it over to the research team once they arrived at GSTT for their 3-month conventional radiograph exam. Alternatively, patients filled out weekly digital diaries emailed by the research team.

All participant's GPs were contacted to collect any scaphoid-related NHS resource use data, (e.g. GP appointment, nurse appointment, telephone appointment, physiotherapy appointments, secondary care appointments) up to a period of 6 months following recruitment.

Table 8. Study flowchart for both control and intervention groups (presented in chronological order)

Activity	**Timing of activity**						
	Prior to Registration for study	**Immediately after registration into the study**	**During MRI**	**After the ED episode** (months)			
				1	**3**	**6**	**Responsible**
Pre-registration evaluation by trained clinician and assessment of eligibility (following an initial x-ray with negative findings for scaphoid fracture)	X						GSTT
Give patient information sheet, explain study and obtain signed informed consent	X						GSTT
Register patient into the study	X						GSTT
Baseline EQ-5D-5L questionnaire		X					GSTT / Participant
Register demographics (e.g. age, gender, previous scaphoid injury, professional occupation) and mechanism of injury		X					GSTT
Give patient registration pack (patient information sheet, copy of consent)		X					GSTT
Patient is randomised into one group		X					GSTT
Discharge patient with supportive splint from the ED (control group) or Book and refer to MRI scan (intervention arm)		X					GSTT
MRI exam and report			X				GSTT
Post EQ-5D-5L questionnaires				X	X	X	GSTT / Participant
Patient experience questionnaires					X		GSTT / Participant
Patient resource diary		X		X	X		GSTT / Participant
3-month research radiographs and research appointment					X		GSTT/ Participant
Retrieve data from GPs						X	GSTT/GPs
Retrieve data from GSTT databases						X	GSTT

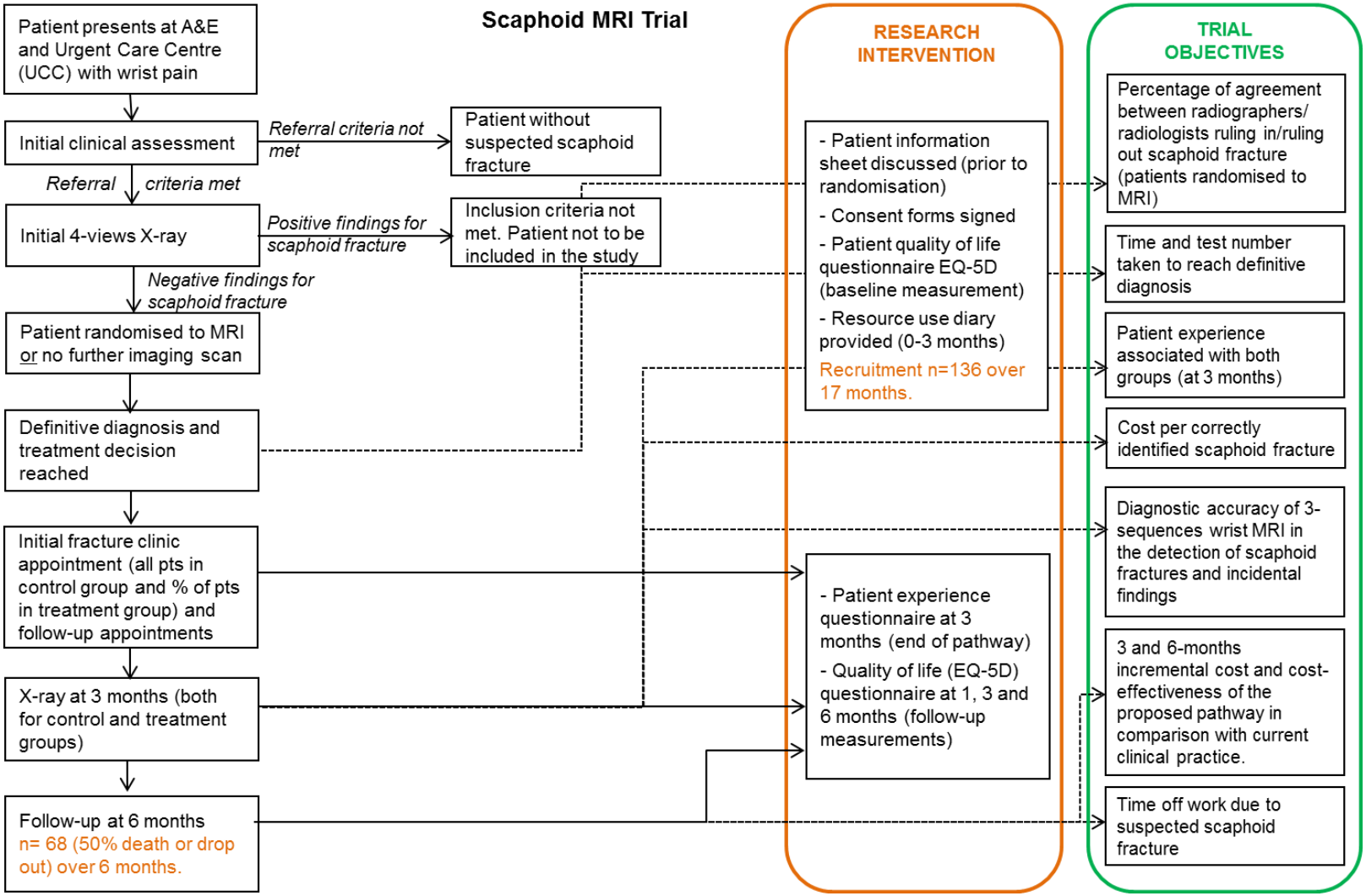


Figure 32. SMaRT trial procedures and respective trial objectives.

Data collection and outcomes

Data were collected by a research team member at baseline and then at 1, 3 and 6 months following recruitment. Data at baseline was collected face-to-face during the initial acute episode whilst follow-up data were collected as per the participant's preference, either via phone, email or post. All data were collected using an electronic case report form (RedCap) and subsequently exported to Microsoft Excel 2013 and statistical analyses were conducted on Stata 15.0.

Participant Demographics:

A variety of information was captured at baseline as part of the SMaRT trial, including:

1. Date of birth (dd/mm/yyyy);
2. Gender (male/female);
3. Ethnicity (e.g. White British, White Other);
4. Postcode (e.g. SE1, SE4);
5. Employment status (e.g. full-time job, part-time job, wholly retired from work);
6. Previous scaphoid fracture (yes/no);
7. Mechanism of injury (e.g. fall on outstretched hand, other injury).

Primary Outcome:

The primary outcome was the 3-month cost difference between: (i) the control group, with radiographs as the only imaging modality in the ED; or (ii) the intervention group, with MRI as an add-on test in the ED for patients with negative findings on the initial radiographs.

Perspective of Analysis

The trial took a NHS and Personal Social Services analytical perspective. Only costs of scaphoid-related NHS diagnostic and treatment events were considered. This approach is consistent with the methodology recommended by the NICE for the evaluation of interventions with potential impact on health outcomes (NICE, 2013). A broader societal perspective was also considered in a secondary analysis.

The estimate of the total costs from a NHS perspective relied on the multiplication of any scaphoid-related healthcare events by the unit cost of each event.

Resource Use Measurement

Resource use data included contacts with an NHS healthcare provider associated with the management of suspected or confirmed scaphoid fractures. A more comprehensive approach to the NHS resource use included data from the following sources of information (visually depicted in Figure 33) based on the participant's NHS number (unique identifier).

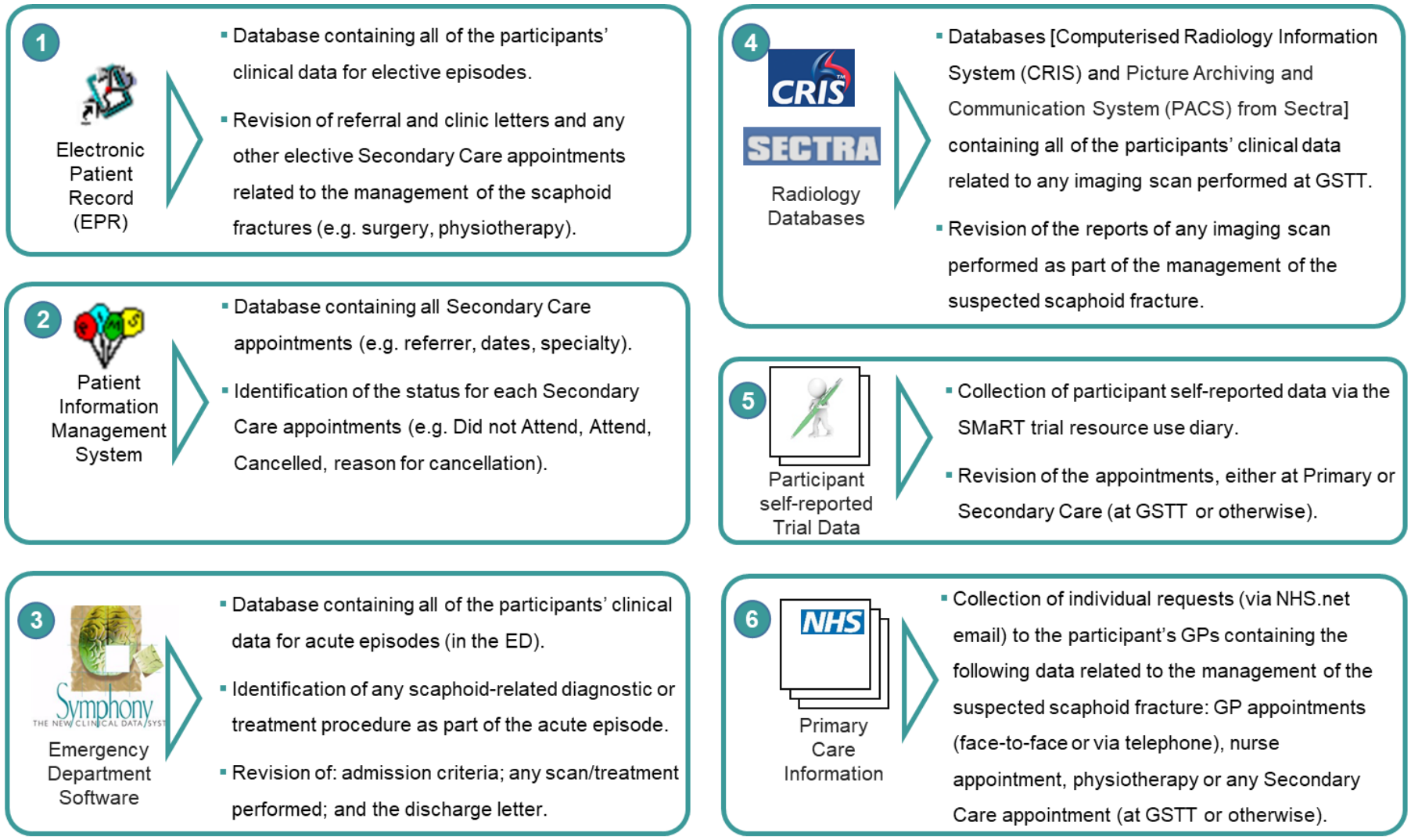


Figure 33. Sources of data merged to measure total NHS resources used in the management of the suspected scaphoid fracture.

1. Electronic Patient Record (EPR). A secondary care database used to maintain the patients' electronic record, with particular focus on clinical data. This database was used as the primary source to individually map the elective diagnostic and treatment pathway timeline associated with the management of the suspected scaphoid fractures. This information allowed us to determine whether a given healthcare contact was scaphoid-related or not.
2. Patient Information Management System (PIMS). A secondary care database to keep the records of referrals, outpatient appointments and inpatient episodes. This database was used as a secondary source to confirm NHS resource use. Furthermore, PIMS provided the status of each appointment (cancelled, did not attend, attend). This information was used to generate costs associated with 'did not attend' events.
3. Symphony (ED software). A secondary care databased used by the ED. This database was used as the primary source to individually map the diagnostic and treatment pathway timeline associated with the acute management of the suspected scaphoid fractures.
4. Computerised Radiology Information System and Picture Archiving and Communication System (Radiology software). These two secondary care databases collected all data (e.g. images and reports) in relation to any diagnostic examination (e.g. radiographs, CT, MRI) performed at GSTT.
5. Participant self-reported trial data. Participants were asked to complete a weekly participant diary and record any visit to any healthcare organisation. This information supplemented the information gathered via secondary care databases, particularly any activity that happened within primary care or secondary care providers other than GSTT.
6. Primary care information. All participants' GPs were contacted to obtain key information related to the management of the suspected scaphoid fracture, including: GP appointments (face-to-face or via telephone), nurse appointment, physiotherapy or any secondary care appointment (at GSTT or other hospitals). This information was used as the primary source to individually map the pathway that happens outside GSTT's remit (e.g. derive the number of GP appointments in the first three months following recruitment, appointments at other Hospitals).

Although marginal, medication costs were mainly associated with painkillers (e.g. paracetamol, ibuprofen) bought over the counter. Given that any out-of-pocket cost were not considered from the NHS payer's perspective, medication data were not included in the present study.

Valuation of Unit Costs

For the purposes of the primary outcome, the valuation of unit costs was, whenever possible, based on NHS Reference Costs 2016-17 (Department of Health and Social Care, 2016). All secondary care contacts were costed using this strategy with the exception of when no tariff was available (e.g. no specific tariff for the intervention, the abbreviated wrist MRI scan). It is important to highlight

that NHS Reference Costs include all costs associated with the provision of care and respective allocation of overheads.

Table 9 lists all the unit costs considered to estimate the primary outcome, including the reference and a brief rationale or any assumptions. Some healthcare events are costed such as inpatient episodes, whilst others, like the provision of a CT or MRI, are presented as individual unbundled tariffs. The use of National Reference Costs, rather than individual costs from GSTT, was considered to enhance the generalisability of the trial's findings. In fact, given that hospital unit costs (e.g. cost of individual MRI) are known to be highly variable across different NHS Trusts (Glick et al. 2007), the use of NHS Reference Costs allows for a better understanding of the cost distribution from the NHS payer's perspective. However, the utilisation of hospital charges, rather than actual costs of provision of care, might not reflect opportunity costs and could lead to different findings (Drummond et al. 2004; Glick et al. 2007). To explore this effect, a sensitivity analysis was considered to evaluate whether the use of existing reimbursement strategies and hospital costs would lead to any difference in the cost analyses.

For primary care contacts, an average cost for appointment (e.g. GP face-to-face appointment, GP phone appointment) was derived from the Unit Costs of Health and Social Care 2016 (Curtis and Burns, 2016) and then inflated to 2017 using the hospital & community health services index. The average GP face-to-face, phone consultation and home visits were estimated to be 9.2, 7.1 and 23.4 minutes long (Curtis 2013), respectively, i.e. equivalent to £36.47, £14.80 and £118.10. Nurse face-to-face and phone appointments were assumed to have the same ratio as the GP appointments, with an estimated cost of £19.50 and £8.00, respectively.

Two additional assumptions were considered in the valuation of unit costs. First, in the case of a 'Did Not Attend (DNA)' event, it was considered that the NHS still incurred in a cost, assumed to be equivalent to 50% of the unit cost where the participant had attended. Given its potential importance, this assumption was subjected to deterministic sensitivity analysis to better understand of its impact on the primary outcome. Second, in the case of an interrupted MRI scan due to an unforeseen claustrophobia event (i.e. no images acquired and no written report), a proportion of the 25% original unit cost was estimated. This estimate was based on the time it took to position and remove the participant from the actual MRI scanner (i.e. the amount of time the scanner was unavailable as no report ever took place).

Table 9. Unit costs for all primary and secondary care events considered in the SMaRT trial.

Category	Unit Type	Unit cost (£)	Reference
Primary care			
GP appointment (face-to-face)	Per appointment	£36.47	Unit Costs of Health and Social Care (Curtis and Burns 2017) and inflated to 2017 using the hospital & community health services index
GP home visit	Per appointment	£118.10	
GP phone appointment	Per appointment	£14.80	
Nurse appointment (face-to-face)	Per appointment	£19.50	
Nurse phone appointment	Per appointment	£8.00	
Secondary care			
Short-sequence wrist MRI scan in the ED	Per scan	£72.40	The immediate acute wrist MRI scan took on average 15 minutes (as opposed to 25 minutes for the full wrist MRI scan). The short-sequence wrist MRI was estimated as a proportion of the cost of the full wrist MRI (15/25*£120=£72.40).
ED episode	Per episode	£73.00	Emergency episode codes (BB11Z - Emergency Medicine, No Investigation with No Significant Treatment). Reference Costs 2016/17 (Department of Health and Social Care 2016). RD01A: Magnetic Resonance Imaging Scan of One Area, without Contrast, 19 years and over; RD20A - Computerised Tomography Scan of One Area, without Contrast, 19 years and over. Outpatient appointments (either at Fracture Clinic, Physiotherapy or occupational therapy) and radiographs were based on local hospital costs (no specific HRG code on NHS reference costs).
Initial / Follow-up Fracture Clinic appointment	Per appointment	£165.00 / £98.00	
Conventional radiograph (scaphoid view) – 1 wrist / 2 wrists	Per scan	£18.30 / £25.80	
Wrist CT	Per scan	£59.60	
Wrist MRI (full sequence)	Per scan	£120.70	
Physiotherapy – First / Follow-up appointment	Per appointment	£54.00 / £45.00	
Occupational therapy – First / Follow-up appointment	Per appointment	£54.00 / £45.00	
Surgeries	Per procedure	£815 to £4,998	Reference Costs 2016/17 (Department of Health and Social Care 2016) – codes HB55C, HR06A, HN42B and HT43C
Plaster cast	Per cast	£36.00	Internal GSTT costing data.
Wrist splint	Per splint	£17.00	

Secondary Outcomes:

This subsection summarises the additional data collection methods used to gather data on the seven secondary outcomes.

I. To perform a cost analysis at 6 months associated with two clinical pathways in the Emergency Department (ED).

This outcome considers the extension of the base case cost analysis up to the 6-months period using the same principles of resource measurement and cost valuation.

II. To perform a cost-effectiveness analyses at 3 and 6 months to estimate the incremental cost-effectiveness ratio (ICER) associated with the proposed intervention.

The cost analysis was performed as per the principles detailed for the primary outcome (3-month total costs) and the above mentioned secondary outcome (6-month total costs). The incremental analysis of effectiveness considered two measures of effect: (i) pain/discomfort levels reported by participants in the resource use diary; and (ii) QALYs, a generic measure of quality of life (i.e. a cost-utility analysis). QALYs were estimated from utility scores derived from the EQ-5D-5L questionnaire (Devlin et al. 2018) at four points in time: baseline (month 0), 1, 3 and 6 months using area under the curve methods assuming linear movement between adjacent points (Drummond et al. 2004). Cost-utility analysis is the preferred method of economic evaluation according to NICE guidelines (NICE 2011b; 2013).

III. To estimate the mean cost per correctly diagnosed scaphoid fracture in the intervention group compared to the control group.

In this secondary outcome, correctly diagnosed scaphoid fractures (either correctly ruled-in or ruled-out) were considered as the measure of effect. This was estimated based on an incremental cost per correct diagnosis (see Equation 1). The decision on whether a correct diagnosis was made by a comparison with the 3-month wrist radiographs (assumed to be the reference). For instance, if both the initial MRI and the final 3-month radiographs showed no evidence of scaphoid fracture, then it was considered to be a correct diagnosis (in this case rule-out diagnosis). This clinical information was derived from the CRIS database, where all medical imaging reports are stored.

Equation 1. Estimate of the cost difference per correct scaphoid diagnosis.

$$\Delta cost\ per\ correct\ diagnosis = \frac{total\ diag.\ costs\ MRI\ group - total\ diag.\ costs\ Control\ group}{total\ of\ correct\ diag.\ MRI\ group - total\ of\ correct\ diag.\ Control\ group}$$

IV. To assess the overall patient satisfaction associated with the intervention group compared to the control group.

This outcome was assessed via a participant 5-point Likert scale non-validated questionnaire three months post-recruitment.

V. To estimate the accuracy associated with the intervention (i.e. immediate wrist MRI) in the detection of scaphoid fracture compared to the control group.

The accuracy of MRI or conventional radiograph during the ED episode was assessed by dividing the number of correct diagnoses (either true negative or true positive findings) by the total number of assessments (see Equation 2). The decision of whether a correct diagnosis was made by comparison against the 3-month wrist radiographs (assumed to be the reference).

Equation 2. Estimate of the accuracy associated with a diagnostic test.

$$Accuracy = \frac{number\ of\ True\ Negatives + number\ of\ True\ Positives}{total\ of\ asessments}$$

VI. To estimate the mean time taken to reach a definitive diagnosis and the first major treatment decision in the intervention group compared to the control group.

The mean time to reach a definitive diagnosis was measured in days as a result of combining different secondary care databases, particularly: (i) Symphony (ED), for the acute part of the pathway (i.e. during the initial episode); and (ii) Electronic Patient Record (EPR) and (iii) CRIS (Radiology database), for the non-acute part of the pathway. The time taken to reach a definitive diagnosis was considered to be when a scaphoid fracture was effectively ruled-out or ruled-in. For instance, if the immediate MRI correctly ruled-out a scaphoid fracture then that would be the time a definitive diagnosis was reached. In contrast, if the findings on the initial MRI were found to be inaccurate (e.g. subsequent imaging found a scaphoid fracture), the latter time point was considered to be the definitive diagnosis.

The mean time spent in the ED in both groups was also evaluated. This variable was measured by combining information from different databases, particularly ED (software called Symphony) and radiology databases. This variable was relevant to plan operational changes and assess potential challenges associated with the acute provision of immediate MRI for these patients.

VII. To estimate the mean time off work or informal care needs due to the suspected scaphoid fracture in the intervention group compared to the control group.

Participants were asked to record the amount of time off work and informal care in a weekly scaphoid diary. These self-reported data were then used to evaluate potential differences between groups in terms of days off work and informal care and estimate potential implications from a societal perspective (as part of a secondary analysis).

3.2.7 Statistical Analyses

Analysis Population:

All analyses were based on the principle of intention-to-treat (ITT), i.e. all participants were analysed as per the allocated group, regardless of whether they actually received the intended treatment, any protocol deviations or potential losses to follow-up (Gupta 2011). This is in contrast to a per-protocol analysis in which only participants that completed the study without any major

protocol deviations are included in the analysis (Gupta 2011; Ranganathan, Pramesh, and Aggarwal 2016). The ITT analysis reflected the pragmatic design of the SMaRT trial, aimed at accepting real-life clinical practice with its non-compliance and protocol deviations and is recommended by the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) (Ramsey et al. 2005). Additionally, by not excluding non-compliers, statistical power was preserved. In summary, the ITT analysis aimed to minimise any potential analytical bias and was performed in accordance to the Consolidated Standards of Reporting Trials (CONSORT) checklist (Consort 2010).

Data Cleaning and Data Validation:

All baseline and follow-up data cleaning were performed prior to data analysis.

Baseline data were captured via a paper-based Case Report Form (CRF) during the ED episode and then added to a web-based CRF. The web-based CRF, designed prior to the start of the trial, specified data edit checks (e.g. time of MRI scan prior to the time of randomisation), hence preventing major data issues. Furthermore, during this process, the PhD student screened the data looking for inconsistencies. In the presence of any potential data errors in the original hand-written data packs, participants or members of the research team were asked for clarification (e.g. date of birth and age do not match) and amendments were made to the original dataset (data editing).

The NHS resource use considered in the primary outcome was derived from the merges of six different sources of information, grouped in two areas: (1) medical records; and (2) self-reported data from participants. These comprehensive data collection methods (detailed in subsection 3.2.6. Data collection and outcomes) included the validation of data using multiple datasets. Potential data errors or inconsistencies were automatically flagged and subsequently edited in the SMaRT trial database. For example, elective secondary care resource use was primarily evaluated using information from EPR. Data were then cross-referenced with other datasets, such as: PIMS; CRIS; and GP information (for other Hospitals than GSTT). Any data inconsistency was reviewed at by the student and edited in the main SMaRT trial database.

Missing Data:

Participants were not excluded from the analysis due to missing data, particularly data related to the primary outcome. In fact, only data from participants that withdrew the informed consent was not considered in the analyses. Data from primary care was considered to be missing completely at random as it related to the participant's GP rather than any participant or disease-related event (Gray et al. 2011). Where data could not be retrieved (e.g. missing primary care data), mean values from the respective group were imputed.

Baseline comparability of randomised groups:

Continuous data were summarised by frequency, mean, standard deviation, minimum, first and third quartile, median and maximum. Tabulations of frequencies for categorical data were presented, as well as the percentage (%) relative to number of non-missing values within the

respective intervention group, unless otherwise specified. Numbers of missing values were reported for both continuous and categorical variables.

The baseline variables described were: age; gender; previous scaphoid fracture (Yes/No); professional occupation; mechanism of injury; baseline EQ-5D-5L; and number of wrists with suspected scaphoid fracture (one/two; stratification variable). No significance testing was performed on the baseline variables between groups given the randomised design of the SMaRT trial (de Boer et al. 2015; Peacock, Kerry, and Balise 2017).

Primary and secondary outcomes: statistical analyses

Differences in mean outcomes between patients randomised to the control (no MRI) and the intervention (MRI) were analysed on an ITT approach. The cost analyses used generalised linear models (GLM) to model the outcome, using an appropriate distribution family. The option for a GLM with a gamma distribution to model the NHS costs was due to the non-negative and positively skewed distribution of costs, with a few patients responsible for very high costs. An identity link function instead of a log link was considered in order to avoid potential biases (Polgreen and Brooks 2012). GLM was used to model the mean cost directly rather than transformation methods (Peacock, Kerry, and Balise 2017). In addition, bootstrap models were considered for the analysis (Gray et al. 2011) and compared against the GLM analysis. Gray et al. (2011) recommends the use of both GLMs and bootstrapping techniques for dealing with skewed cost data. Group difference estimates and associated confidence intervals were reported, together with p-values.

Given the short time period (<1 year) associated with the management of suspected or confirmed scaphoid fractures, it was not deemed appropriate to consider discounting of either costs or effects.

<u>Primary Outcome</u>

The primary outcome of 3-month cost data were analysed using a GLM, with the following predictors used in the model:

- Randomisation group (intervention group/control group [main predictor]);
- Number of wrists with suspected fractures (one/two [stratification variable]).

<u>Secondary Outcomes</u>

Other cost analyses were conducted using a GLM and bootstrap analyses using the intervention group as predictor. Cost-effectiveness analyses at 3 and 6 months were performed using bootstrap models using pain scores and QALYs as the measures of effect. If utility data were missing, multiple imputation methods were used based on the assumption that the data were missing at random. Missing data were imputed using ‘multiple imputation using chained equations’ (MICE), with the number of multiply imputed datasets to be equal to the fraction of incomplete service-use information (White, Royston, and Wood 2011). 1000-replicate bootstrap analyses showing difference in costs and outcomes were presented on cost-effectiveness planes. Cost-effectiveness acceptability curves showing the probability that intervention was more cost-effective than control

at varying thresholds of willingness to pay were presented. All analyses were performed using the software Stata 15.0 for Windows (StataCorp LLC, Lakeway Drive,Texas).

NHS resource use was also evaluated. Depending on the normality assessment using the Shapiro-Wilk test, the use of independent t-test or Mann-Whitney U tests were used to test for differences in mean utilisation between groups.

Differences in clinical findings between the two groups were, given the dichotomous nature of the variable (presence/absence of injury), evaluated using the Pearson Chi-square statistical test.

Patient satisfaction in both groups was assessed based on a 3-month non-validated patient satisfaction questionnaire (5-point Likert scale). The use of Chi-square test was used to test the hypothesis that there was no differences between groups.

The time elapsed in the ED and time taken to reach a definitive diagnosis and time taken off work and informal care in each group was also evaluated. In both variables, depending on the normality assessment using the Shapiro-Wilk test, the use of independent t-test or Mann-Whitney U tests were used to test for differences between groups.

Sensitivity Analyses

Several deterministic sensitivity analyses around key model parameters were performed. The NHS resource use was retrieved from the trial but the unit cost valuation was based on national data (e.g. NHS reference costs) or, if not feasible, unit costs from the GSTT's perspective or assumptions. Given potential impact on the trial's primary outcome, the following unit costs were subjected to deterministic sensitivity analyses: (i) immediate wrist MRI in the ED; (ii) fracture clinic appointments (both first and follow-up outpatient appointments); (iii) the 'Did not attend' and interrupted MRI events; (iv) all events using existing reimbursement strategies as a proxy of cost. Finally, an additional cost analysis took a broader societal perspective of analysis.

3.3 Results

3.3.1 Data Validation and Completeness

Baseline data were 100% complete, with the exception of data related to the ED timeline (e.g. time of MRI scan). In the latter case, missing data were retrospectively collected via secondary care databases (ED Symphony and CRIS). These data were used to estimate one of the secondary outcomes, time taken to reach a definitive diagnosis or time elapsed in the ED.

Follow-up data were captured by combining data from primary and secondary care databases and self-reported data from participants. Data from secondary care databases were 100% complete. Data from primary care databases were 98.5% complete (n=130) and self-reported data were 55% (n=72) and 53% (n=70) complete at 3 and 6 months post-recruitment, respectively. In the absence of data from both the primary care databases and self-reported data, any resource use outside GSTT was missing. Missing values for primary care utilisation were imputed using the mean values from the respective group. However, this imputation was unlikely to impact the results given: (i) the

completeness of data from primary and secondary care databases; and (ii) the low level of use of primary care resources.

3.3.2 Participant Flow

Participant flow associated with the SMaRT trial is illustrated in the CONSORT diagram (Figure 34). 313 patients were assessed for eligibility over a 21 months period in the ED. Out of these, 43.5% of patients were recruited (n=136) and the remaining 56.5% (n=177) were not recruited. For those not recruited, the reasons were as follows: 57.6% of participants did not want to participate (n=102), 29.4% of participants did not meet the inclusion criteria (n=52), 8.5% lived outside the catchment area and did not want to be followed-up at GSTT (n=15) and 4.5% had significant language barriers (n=8).

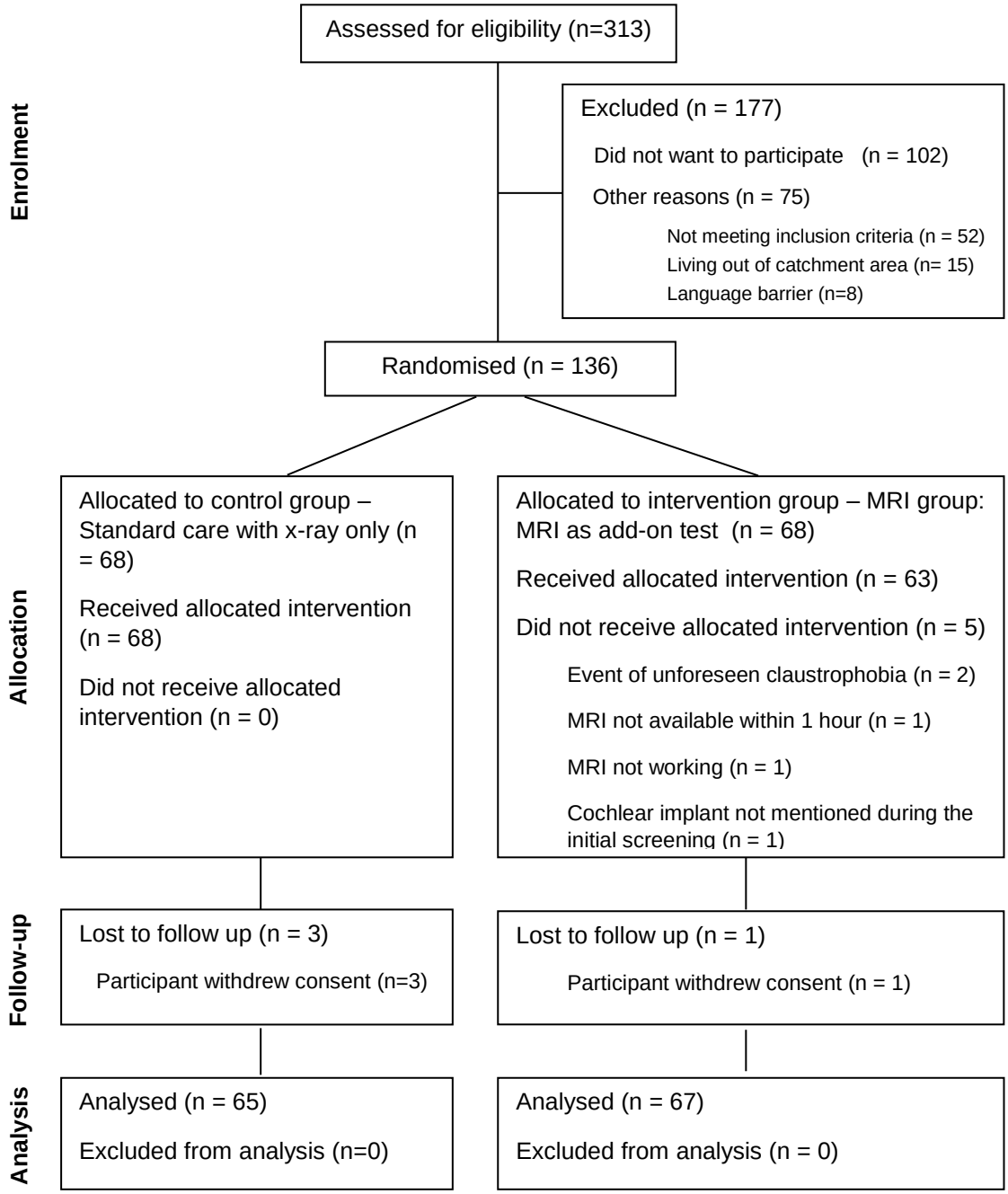


Figure 34. Participant flow chart for the SMaRT trial.

A total of 136 participants were equally randomised to the control group (standard care, n=68) and the intervention group (MRI group, n=68). One 100% (n=68) and 92.6% (n=63) of the participants randomised, respectively, to the control and intervention groups received the allocated treatment. Five participants randomised to the intervention group did not receive the allocated treatment for the following circumstances (see Figure 34): unforeseen claustrophobia (n = 2); MRI not available within 1 hour (n = 1); MRI not working (n = 1); and cochlear implant not mentioned during the initial screening (n = 1).

With regards to the follow-up period, 4.4% (n=3) and 1.5% (n=1) of participants withdrew the informed consent in the control and intervention group, respectively, and were considered lost to follow-up. No further participants were lost to follow-up. All participants who did not withdraw their informed consent were included in the analysis, equivalent to 95.6% (n=65) and 98.5% (n=67) participants in the control and intervention groups, respectively.

3.3.3 Participant Characteristics – Baseline

Table 10 (for categorical variables) and Table 11 (for quantitative variables) summarise the baseline sociodemographic and baseline outcome variables by randomisation group. All participants, apart from four who withdrew consent, were included in the baseline analysis (n=132), distributed in the control group (n=65) and the intervention group (n=67). There were no missing data, apart from some observations in the 'mechanism of injury' variable (detailed below). No significance testing between groups was performed given the trial's randomised design.

Categorical data:

Table 10 describes the participant's characteristics (gender and employment status) and the wrist injury (previous scaphoid injury, mechanism of injury, dominant hand and arm injured) organised by randomisation group.

The majority of participants were male, both in the control group (52%) and the intervention group (61%). With regards to employment status, the majority of participants were employed full-time (over three quarters in both groups), employed in a part-time job (9% and 3% in the control and intervention group, respectively) or self-employed (3% and 6% in the control and intervention group, respectively).

The majority of participants (85% and 84% in the control and intervention group, respectively) had no previous scaphoid injury. With regards to the current scaphoid injury, 55% of the participants identified falling on an outstretched hand as the mechanism of injury. The majority of participants were right-handed (89% and 93% in the control and intervention group, respectively). The arm injured during the present episode was predominantly the left arm in the control group (51%) compared to the right arm in the intervention group (57%). Only three participants in the control (n=1) and intervention group (n=2), respectively, had both wrists injured.

Continuous data:

The average age (SD) of participants was 36.2 (12.6) and 38.2 (13.4) years in the control and intervention group, respectively (Table 11).

When considering the utility at baseline, as measured using the questionnaire EQ-5D-5L, the mean utility value (SD) was 0.786 (0.158) and 0.822 (0.139) for the control and intervention group, respectively. Similarly, when considering the self-reported health score, participants in the intervention group scored higher, with a mean score (SD) of 78.7 (13.8) compared 72.1 (16.8) in the control group. It is relevant to note that the EQ-5D-5L questionnaire was applied prior to randomisation in order to avoid potential biases.

Table 10. Descriptive statistics of the categorical variables at baseline.

		Randomisation Group			
		Control group (n=65)		Intervention group (n=67)	
		N	%	N	%
Gender	Male	34	52%	41	61%
Employment Status	Employee in full - time job (30 hours or more/week)	51	78%	53	79%
	Employee in part - time job (under 30 hours/week)	6	9.2%	2	3.0%
	Self-employed, full or part - time	2	3.1%	4	6.0%
	Full - time education at school, college or university	0	0.0%	3	4.5%
	Looking after the home	0	0.0%	1	1.5%
	Wholly retired from work	1	1.5%	1	1.5%
	Unemployed and available for work	3	4.6%	1	1.5%
	Permanently sick/ disabled	2	3.1%	1	1.5%
	Doing something else	0	0.0%	1	1.5%
Previous Scaphoid Injury	Yes	10	15%	11	16%
Mechanism of Injury	Fall on an outstretched hand	36	55%	37	55%
	Other injury	29	45%	30	45%
Dominant hand	Left	7	11%	5	7.5%
	Right	58	89%	62	93%
Arm Injured	Left	33	51%	27	40%
	Right	31	48%	38	57%
	Both	1	1.5%	2	3.0%

Table 11. Descriptive statistics of the three numerical variables at baseline per randomisation group: age; utility at baseline; and overall visual analogue score (VAS) at baseline (estimated from EQ-5D-5L questionnaire).

			Count (N)	Mean	Standard Deviation	Minimum	Percentile 25	Median	Percentile 75	Maximum
Age	Randomisation group	Control	65	36.2	12.6	18.0	27.0	32.0	42.0	73.0
		MRI	67	38.2	13.4	20.0	27.0	36.0	46.0	71.0
Utility at baseline	Randomisation group	Control	65	0.786	0.158	0.330	0.732	0.825	0.893	1.000
		MRI	67	0.822	0.139	0.273	0.747	0.837	0.927	1.000
Visual analogue score at baseline	Randomisation group	Control	65	72.1	16.8	30.0	60.0	75.0	85.0	100.0
		MRI	67	78.7	13.8	40.0	70.0	80.0	90.0	100.0

3.3.4 Clinical findings

All recruited participants had negative findings on the initial scaphoid radiographs. Nonetheless, given its limited ability to rule-out fractures on presentation, potential injuries, such as scaphoid or other bone fractures, might subsequently be diagnosed a few days or weeks further down the line.

Scaphoid fractures:

A total of 11 (8.3%) scaphoid fractures were diagnosed, with 7 (10.4%) in the intervention group (i.e. MRI group) and 4 (6.2%) in the control group. Table 12 expands on the type of scaphoid injury and the time of diagnosis for each injury. For the intervention group, all fractures were detected on presentation to the ED, whilst for the control group, scaphoid fractures were detected from 8 to 48 days after presentation to the ED. One participant in each group underwent surgical treatment for confirmed scaphoid fractures. Figure 35 illustrates three views from the immediate MRI performed in the ED where a scaphoid waist fracture is visible (despite the initial negative radiographs).

Given the dichotomous variable, the Pearson Chi-square statistical test was used. The differences in the proportion of scaphoid fractures between the two groups were not deemed statistical significant (p=0.372).

Table 12. Number and type of scaphoid fractures diagnosed in both groups.

	Scaphoid fractures		
	Control group (n=65)		Intervention group (n=67)
1.	Subtle undisplaced fracture line noted on one view around the waist of the scaphoid.	Found on repeated x-ray 8 days after ED visit	Trabecular scaphoid and undisplaced distal pole fractures.
2.	Undisplaced or very minimally displaced fracture of the proximal pole of the scaphoid.	Found on MRI 48 days after ED visit	Undisplaced fracture through the waist of the scaphoid
3.	Undisplaced fracture of the scaphoid waist	Found on MRI 9 days after ED visit	Incomplete scaphoid fracture.
4.	Closed fracture of the scaphoid waist.	Found on repeated x-ray 13 days after ED visit	Scaphoid tubercle fracture.
5.			Impaction fracture of the radial cortex of the scaphoid, though no evidence of scaphoid waist displacement.
6.			Impaction fracture of the scaphoid tubercle with no evidence of displacement.
7.			Undisplaced fracture of the body of the scaphoid.

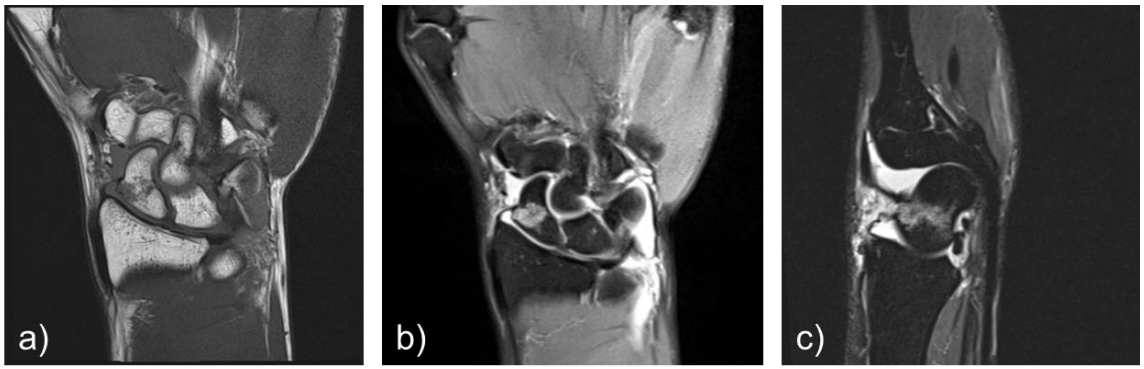


Figure 35. Imaging of patient with fracture of the scaphoid waist showing the abbreviated MRI: (a) coronal T1; (b) coronal PDFS; and (c) sagittal STIR views.

Other fractures:

Similarly to the scaphoid bone, other bone injuries might be present despite the negative findings in the initial radiographs. These injuries could lead to participants re-presenting to secondary care or being diagnosed as part of the follow-up procedures.

A total of 20 (15.2%) other bone fractures (besides scaphoid fractures) were diagnosed, with 15 (22.4%) in the intervention group and 5 (7.7%) in the control group. Table 13 expands on the type of injury and time of diagnosis for each group. All other bone fractures in the intervention group (MRI group) were detected on presentation to the ED except a fracture of the hook of the hamate was found on CT 116 days after the ED visit. In the control group, all five fractures were diagnosed from 8-12 days after the ED presentation.

Given the dichotomous variable, the Pearson Chi-square statistical test was again used. The differences in the proportion of non-scaphoid fractures between the two groups were statistically significant (p=0.019), with the MRI group presenting a higher proportion.

Table 13. Number and type of other bone fractures (besides the scaphoid) diagnosed by groups.

	Other bone fractures		
	Control group (n=65)		Intervention group (n=67)
1.	Right radial styloid fracture	Found on repeated x-ray 9 days after ED visit	Undisplaced distal radial fracture.
2.	Distal radial fracture with no displacement.	Found on CT 12 days after ED visit	Undisplaced triquetrum fracture (with partial lunotriquetral ligament tear)*
3.	Extra-articular undisplaced distal radius fracture.	Found on repeated x-ray 8 days after ED visit	Impaction fracture of the proximal articular surface of the second metacarpal
4.	Undisplaced capitate fracture and contusion of the pisiform and trapezium (from the MRI)	Found on MRI 8 days after ED visit	Distal radial fracture with minimal displacement.
5.	Undisplaced fracture of the distal radius	Found on MRI 11 days after ED visit	Cortical disruption of the articular surface of the distal radius.
6.			Undisplaced fracture of the distal radius with minor intraarticular extension, involving the radial styloid.
7.			Undisplaced distal radial fracture.
8.			Fracture of the hook of the hamate (found on CT 116 days after ED visit)
9.			Subtle base of fifth metacarpal intra-articular fracture with no displacement.
10.			Fracture of the distal radius
11.			Undisplaced intra-articular distal radial fracture and nondisplaced transverse trabecular fracture of the base of the fifth metacarpal
12.			Closed fracture of the capitate
13.			Pisiform fracture
14.			Nondisplaced trabecular fracture of the lunate with TFCC injury. *
15.			Undisplaced fracture of the distal radius.

* patients with concomitant soft tissue injuries also included in Table 14.

Soft tissue injuries:

Finally, major soft tissue injuries can also occur (e.g. ligament rupture). A total of 5 (3.8%) major soft tissue injuries were diagnosed, with 4 (6.0%) in the intervention group and 1 (1.5%) in the control group. Table 14 expands on the type of wrist injury diagnosed in each arm. It is important to note that minor soft tissue injuries (e.g. tissue oedema) were not included in this analysis.

A participant in the control group underwent two surgeries in order to treat a complete scapholunate ligament rupture. Two of the patients in the intervention group in Table 14 with reported soft tissue injuries also had bone injuries and were therefore included in Table 13, meaning that 17 patients in the intervention group had injuries detected other than scaphoid fractures (Figure 36).

The differences in the proportion of other soft tissue detected injuries between the two groups showed a trend of statistical significance (p =0.102).

Table 14. Number and type of soft tissue / ligamentous injuries diagnosed in both groups.

	Soft tissue injuries	
	Control group (n=65)	Intervention group (n=67)
1.	Complete scapholunate ligament injury/rupture (11 days after ED visit)	Undisplaced triquetrum fracture (with partial lunotriquetral ligament tear)*
2.		Partial tear of the ulnar and dorsal ulnolunate ligament at the dorsal edge of the TFCC. Partial TFCC tear and sprain.
4.		Ulnar TFCC tear with a joint effusion extending to the DRUJ and detachment at the meniscocapsular homologue, foveal and ulnar styloid attachments. Probable ECU tendonosis.
5.		TFCC tear (with undisplaced trabecular fracture of the lunate)*

* patients with concomitant soft tissue injuries also appear in Table 13.

Normal findings:

A total of 43 participants (64.2%) randomised to the intervention group had normal findings in the initial wrist MRI (Figure 36). In the control group, as part of the inclusion criteria, all participants had negative findings in the initial conventional radiograph. Out of the 65 participants randomised to the control group, 55 (84.6%) were subsequently found to have normal findings during the diagnostic pathway (including the final 3-month X-ray).

Summary of clinical findings:

Figure 36 summarises the key clinical findings organised per randomisation group. As detailed in the previous subsections, the intervention group was associated with a higher number of clinically relevant findings, both bone fractures and soft tissue injuries.

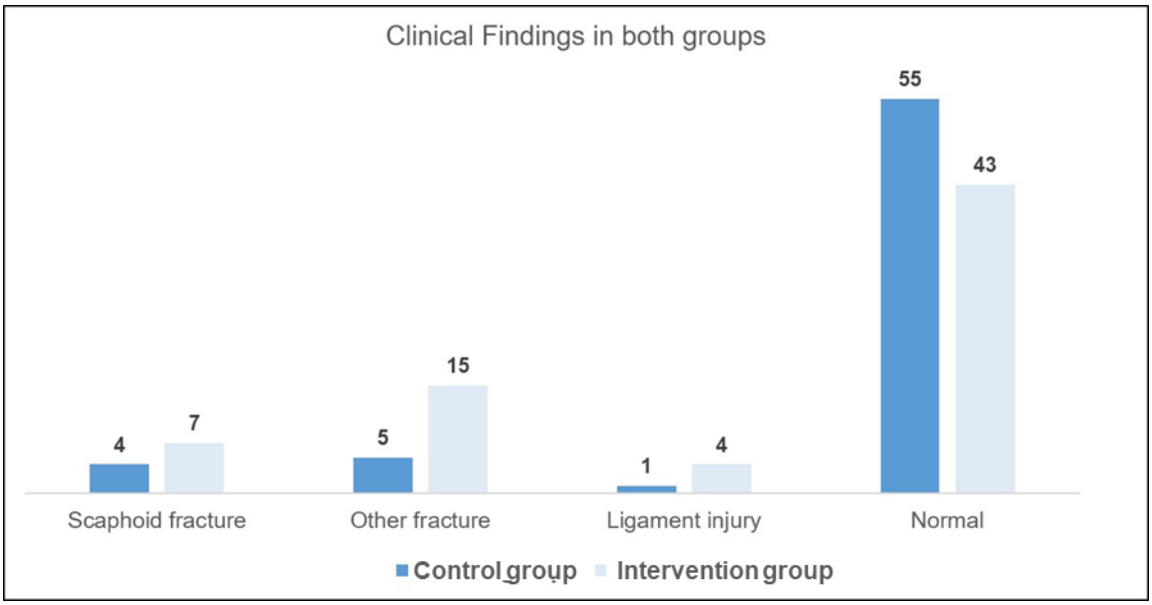


Figure 36. Distribution of clinical findings grouped per randomisation group.

3.3.5 Primary Outcome

Primary Outcome:

The primary outcome was to estimate the 3-month costs associated with both groups. For this purpose, all participants were followed-up for a period of three months to capture all relevant NHS resources used in the management of the suspected scaphoid fracture. This included both primary and secondary care resources.

Table 15 summarises the mean number of NHS events per participant for both arms. A mean of 6.8 and 5.7 NHS episodes were performed in the control (n=65) and intervention groups (n=67) (p=0.105), respectively. With regards to primary care contacts, the control group attended a higher number of appointments with the difference between groups being statistically significant (p=0.02). If all secondary care contacts are considered, the control group presented a slightly higher utilisation but this was not statistically significant (p=0.189).

Table 15. Mean (SD) NHS events per participant per type of healthcare provider for both groups.

		N	Mean	Std. Deviation
Primary care	Control group	65	0.59	1.43
	Intervention group	67	0.18	0.63
Secondary care	Control group	65	6.25	3.72
	Intervention group	67	5.48	3.01
Total NHS contacts	Control group	65	6.83	4.37
	Intervention group	67	5.66	3.11

Table 16 details the utilisation of NHS resources per type of event, both for primary and secondary care, and organised per randomisation group.

Table 16. Breakdown of NHS resource use per type of activity by group.

	Control group (n=65)		Intervention group (n=67)		p-value
Type of NHS appointment	**Total of episodes**	**Mean (SD)**	**Total of episodes**	**Mean (SD)**	
Primary care					
GP face-to-face appointment	23	0.35 (0.91)	8	0.12 (0.41)	0.050
GP home visit	1	0.02 (0.12)	0	0 (0)	0.310
GP non face-to-face appointment	8	0.12 (0.42)	4	0.06 (0.49)	0.052
Nurse home visit	4	0.06 (0.30)	0	0 (0)	0.076
Nurse non face-to-face appointment	2	0.03 (0.17)	0	0 (0)	0.149
Secondary care					
Visits to the ED	72	1.11 (0.44)	72	1.08 (0.27)	0.827
MRI in the ED	0	0 (0)	62	0.93 (0.27)	<0.001
Initial Fracture Clinic appointment	61	0.94 (0.53)	32	0.48 (0.50)	<0.001
Follow-up Fracture Clinic appointment	49	0.75 (1.13)	23	0.34 (0.59)	0.022
Conventional radiograph	68	1.05 (0.84)	19	0.28 (0.55)	<0.001
CT	13	0.20 (0.40)	6	0.09 (0.29)	0.073
MRI	17	0.26 (0.44)	0	0 (0)	<0.001
Physiotherapy / Occupational therapy	39	0.60 (1.30)	62	0.93 (1.56)	0.124
Wrist surgery (one patient might have multiple surgeries)	5	0.08 (0.37)	1	0.02 (0.12)	0.290
Splint	68	1.05 (0.21)	70	1.05 (0.27)	0.987
Plaster Cast	14	0.22 (0.52)	20	0.30 (0.60)	0.415

The control group presents a higher number of appointments for all primary care, with the intervention group reporting a statistically significant lower number of GP face-to-face appointments (mean 0.12 vs 0.35, p=0.050).

In relation to secondary care, both groups had similar number of visits to the ED (p=0.827) and, when considering elective appointments, the intervention group had significantly fewer fracture clinic appointments, both first appointments (mean 0.48 vs 0.94, p<0.001) and follow-up appointments (mean 0.34 vs 0.75, p=0.022). Similar to outpatient appointments, the utilisation of radiographs was lower in the intervention group (mean 0.28 vs 1.05, p<0.001). The difference in terms of the use of advanced imaging after the initial acute episode was also evaluated. The intervention group was associated with a lower utilisation of advanced imaging following ED presentation, both for CT (mean 0.09 vs 0.20, p=0.072) and MRI scans (mean 0.0 vs 0.26, p<0.001) (Table 16).

Figure 37 illustrates the high-level follow-up pathway per participant for both randomisation groups. Out of patients randomised to the control group, only a small proportion of 7.7% participants (n=5) had no formal secondary care follow-up. These were participants that either left the ED prior to having a formal Fracture Clinic appointment or did not attend (DNA) secondary care appointments. In contrast, 52% (n=35) of participants in the intervention group had negative MRI findings and hence no secondary care follow-up.

Almost 90% (n=58) of participants in the control group had formal follow-up at secondary care with imaging follow-up, either with radiographs only (49%) or advanced imaging (40%). This 90% figure contrasted with a secondary follow-up of 46% (n=31) in the intervention group. Out of these 31 intervention group participants, 87% (n=27) of the imaging follow-up was based on radiographs only. Four participants (6.0%) had subsequent follow-up with CT as a secondary imaging technique to further analyse bone displacement visualised on the initial MRI exam.

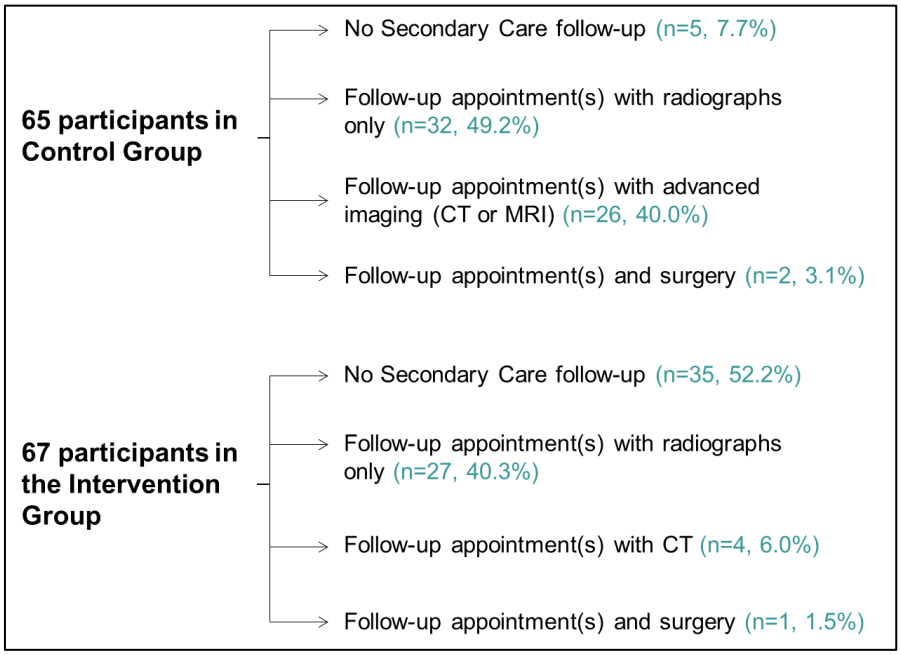


Figure 37. High-level follow-up pathway per participants in both randomisation groups.

Table 17 summarises the 3-month cost distribution (mean, standard deviation, minimum, median, percentiles 25 and 75 and maximum) per randomisation group. As previously detailed, these costs were estimated based on the unit costs (see Table 9) multiplied by the NHS resource use (see Table 16). The mean cost of management per participant [mean (SD)] was higher in the control group compared to the intervention group [£542.40 (£855.20) vs £368.40 (£338.60)], leading to a mean cost difference between groups of £174 per participant. Secondary care costs accounted for 91.4% and 96.8% of the total mean costs of management in the control and intervention groups, respectively.

As expected with cost variables, the cost distribution is positively skewed (mean >> median), as it is affected by a small proportion of patients that have significantly higher costs (maximum cost of £7,116 and £2,691 for the control and intervention group, respectively). This is summarised in Table 17 and illustrated in the histograms in Figure 38 and Figure 39.

Table 17. Descriptive statistics of the three month costs associated with the control (n=65) and intervention groups (n=67).

	3-month cost distribution						
	Mean	Standard Deviation	Minimum	Percentile 25	Median	Percentile 75	Maximum
Control	£542.40	£855.20	£94.00	£259.00	£457.00	£601.00	£7,116.00
Intervention	£368.40	£338.60	£166.00	£166.00	£284.00	£465.00	£2,691.00

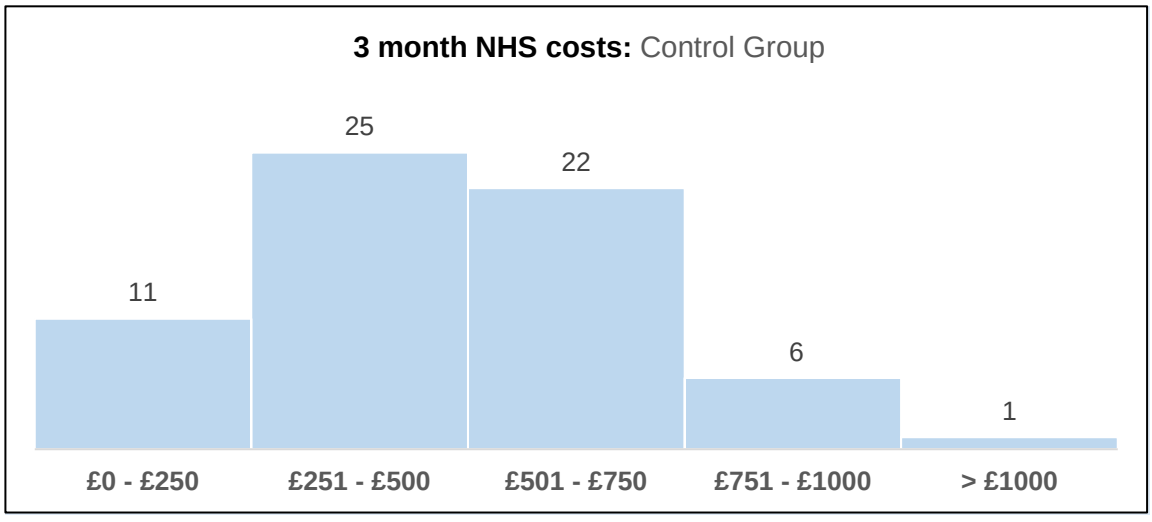


Figure 38. Histogram for the 3-month cost distribution for the control group.

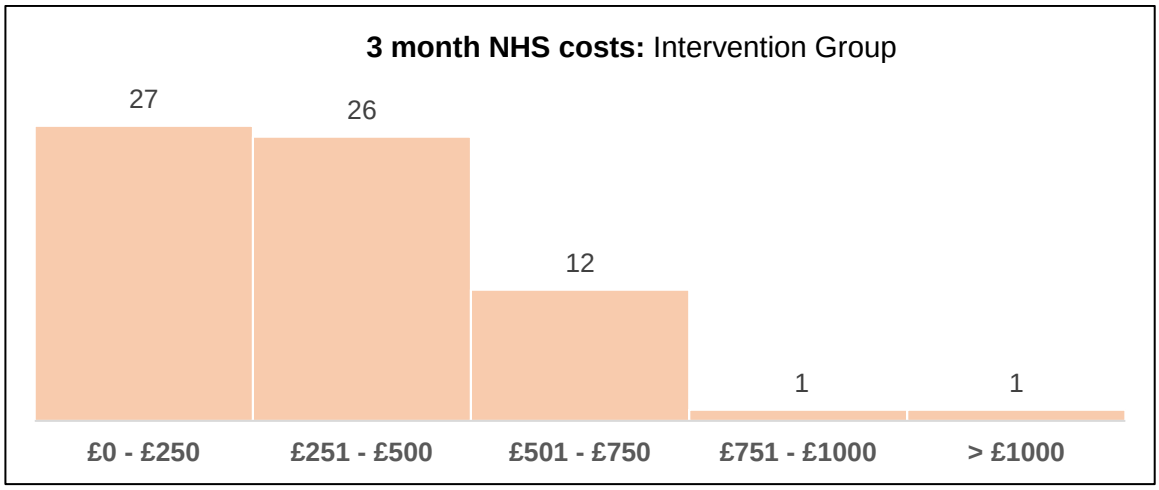


Figure 39. Histogram for the 3-month cost distribution for the intervention group (MRI group).

Taking into consideration the anticipated data skewness, the statistical analysis plan used: (i) GLM to model the 3-month cost analysis used outcome, using an appropriate distribution family, the Gamma family and identity link; and (ii) bootstrap analysis.

The *a priori* Statistical Analysis Plan (SAP) considered the number of wrists with suspected scaphoid fractures as a stratification variable. However, due to the fact that only 3 participants had injured both wrists (2 in the intervention arm and 1 in the control arm) it was not feasible to use this as a stratification variable for the primary outcome.

(i) Generalised Linear Model (GLM)

The results from the GLM analysis (Gamma family and identity link function) for the primary outcome (3-month cost analysis) are detailed in Table 18.

Table 18. GLM analysis for the 3-month cost analysis variable (gamma function, identity link).

	Control group (n=65)	**Intervention group (n=67)**	**Difference (Intervention-Control)**	**95% CI**	**p-value**
Total cost at 3 months	542 (855)	368 (339)	-174	-378 to 30	0.094

The mean cost difference per participant between both groups was -£174 (CI 95%: - £378 to £30, p=0.094). Hence, at 3 months, no statistically significant difference between the two groups was estimated (assuming a p-value of 0.05 as statistical significance).

Additionally, a diagnostic cost was also estimated for both groups. This diagnostic cost includes all NHS costs incurred up to the moment of definitive diagnosis, but exclude any treatment costs, particularly surgical procedures costs. The mean diagnostic cost difference between both groups was estimated at £113 per participant (CI 95%: - £188 to - £39) (Table 19). The 3-month diagnostic costs showed a statistically significant difference between both groups (p=0.003).

Table 19. GLM analysis for the diagnostic cost analysis variable (gamma function, identity link).

	Control group (n=65)	**Intervention group (n=67)**	**Difference (Treatment-Control)**	**95% CI**	**p-value**
Total Diagnostic Costs	454 (242)	341 (187)	-113	-188 to -39	0.003

(ii) Bootstrap analysis

A 1000-replicate bootstrap analysis for the variable total 3 month costs was also performed, grouped by randomisation group. The 95% confidence interval for three types of bootstraps (normal, percentile and bias-corrected) are presented in Table 20. As with the GLM, the same mean difference cost per participant is obtained (- £174) but the 95% confidence intervals varied according to the type of bootstrap analysis considered: normal (CI 95%: - £400 to £52); percentile (CI 95%: - £428 to £12); and bias-corrected (CI 95%: - £453 to -£8). Only the bias-corrected bootstrap analysis showed a statistically significant cost difference per participant (the value 0 is not included in the 95% CI). The use of bias-corrected bootstrap analysis is the most appropriate to assess skewed cost data (Jiang and Zhou 2004).

Table 20. Bootstrap analysis for the variable total cost at 3 months (1,000 replicates).

Variable	Reps	Observed	Bias	Std. Err.	[95% Conf. Interval]		
Randomisation Group	1000	-£174.07	-1.329021	115.13	-£400.00	£51.86	(N)
					-£428.49	£11.76	(P)
					-£453.33	-£7.58	(BC)

Note: N=normal; P=percentile; BC=bias-corrected

3.3.6 Secondary Outcomes

I. To perform a cost analysis at 6 months associated with two clinical pathways in the Emergency Department (ED).

(i) Generalised Linear Model (GLM)

The first secondary outcome considered the time extension of the original cost analysis up to 6 months following the initial ED episode. Table 21 summarises the 6-month cost distribution per randomisation group. The mean cost management per participant (SD) was higher in the control group compared to the intervention group [£660.90 (£1,188.80) vs £395.20 (£344.80)].

Table 21. Descriptive statistics of the six month costs associated with the control (n=65) and intervention group (n=67).

	Mean	Standard Deviation	Minimum	Percentile 25	Median	Percentile 75	Maximum
Control	£660.90	£1,188.80	£94.00	£284.00	£457.00	£666.00	£7,332.00
Intervention	£395.20	£344.80	£166.00	£166.00	£331.00	£491.00	£2,691.00

The GLM analysis (Gamma family and identity link function) for the primary outcome (6-month cost analysis) is detailed below (Table 22).

Table 22. GLM analysis for the 6-month cost analysis variable (gamma function, identity link).

	Control group (n=65)	**Intervention group (n=67)**	**Difference (Intervention-Control)**	**95% CI**	**p-value**
Total cost at 6 months	661 (1,189)	395 (345)	-266	-528 to -3.3	0.047

As with the 3-month cost analysis, the intervention was associated with lower overall cost, with a mean cost difference per participant between groups of £266 (CI 95%: - £528.1 to - £3.3, p= 0.047). Hence, at 6 months, the cost difference between the two groups was statistically significant.

(ii) Bootstrap analysis

As with the primary outcome, a 1000-replicate bootstrap analysis for the variable total 6 month costs was also performed, grouped by randomisation group. The 95% confidence intervals for three types of bootstraps (normal, percentile and bias-corrected) are presented in Table 23. As with the GLM, the same mean difference cost per participant was estimated (-£266) but the 95% confidence intervals varied as per the type of bootstrap analysis: normal (CI 95%: - £558 to £27); percentile (CI 95%: - £580 to -£9); and bias-corrected (CI 95%: - £635 to -£29). Both the percentile and bias-corrected bootstrap analyses showed a statistically significant cost difference per participant (the value 0 is not included in the 95% CI).

Table 23. Bootstrap analysis for the variable total cost at 6 months (1,000 replicates).

Variable	Reps	Observed	Bias	Std. Err.	[95% Conf. Interval]		
Randomisation Group	1000	-£265.68	7.276311	148.99	-£558.05	£26.69	(N)
					-£580.13	-£-9.33	(P)
					-£635.49	-£28.61	(BC)

Note: N=normal; P=percentile; BC=bias-corrected

Outcome II. To perform a cost-effectiveness analyses at 3 and 6-months to estimate the ICER associated with the proposed intervention using QALYs and pain score as the measure of effect.

QALY as the measure of effect: Table 24 summarises the key descriptive statistics (mean, SD, minimum, percentile 25, median, percentile 75, maximum) associated with the utilities and healthcare scores (visual analogue scores), derived from the standardised questionnaire EQ-5D-5L, at four points in time (baseline, month 1, 3 and 6).

The mean utility values at baseline (SD) for the control and intervention group were, respectively, 0.786 (0.158) and 0.822 (0.139) (p=0.166). EQ-5D-5L data completeness decreased during the follow-up period, with 60% and 58% of data being complete at month 3 and 6 post-recruitment, respectively (see Table 24). The mean utility values at both 3 and 6 months post-recruitment were lower in the control group compared to the control group [mean utility at month 3 of 0.924 vs 0.843 (p=0.089) and mean utility at month 6 of 0.950 vs 0.843 (p=0.019)]. Potential imbalances between utilities at baseline need to be taken into account as these were likely to be correlated with the utilities over the follow-up period (Manca, Hawkins and Sculpher 2005).

Table 24. Descriptive statistics for the utility and VAS at baseline and months 1, 3 and 6.

			Mean	Standard Deviation	Minimum	Median	Maximum
BASELINE	Utility	Control (n=65)	.786	.158	.330	.825	1.000
		Intervention (n=67)	.822	.139	.273	.837	1.000
	VAS	Control (n=65)	72.1	16.8	30.0	75.0	100.0
		Intervention (n=67)	78.7	13.8	40.0	80.0	100.0
MONTH 1	Utility	Control (n=42)	.747	.238	.073	.825	1.000
		Intervention (n=50)	.854	.105	.518	.869	1.000
	VAS	Control (n=42)	75.6	17.9	25.0	80.0	95.0
		Intervention (n=50)	81.0	14.9	10.0	84.0	100.0
MONTH 3	Utility	Control (n=36)	.843	.227	-.076	.893	1.000
		Intervention (n=46)	.924	.077	.743	.942	1.000
	VAS	Control (n=36)	82.5	16.5	30.0	88.0	100.0
		Intervention (n=46)	85.2	10.3	50.0	90.0	100.0
MONTH 6	Utility	Control (n=33)	.843	.211	-.076	.893	1.000
		Intervention (n=44)	.950	.068	.709	1.000	1.000
	VAS	Control (n=33)	84.6	9.9	65.0	85.5	100.0
		Intervention (n=44)	89.7	7.4	65.0	90.0	100.0

In order to adjust for imbalances in the mean baseline utilities, a multiple regression analysis was conducted as suggested by Manca, Hawkins, and Sculpher (2005) and detailed below in Equation 3.

Equation 3. Multiple regression analysis with adjustment for baseline utility imbalances.

$$Utility_I = \beta_0 + \beta_1 . t_I + \beta_2 . Q_i^b ,$$

where the index *i* is the participant identifier (i =1, 2,..., 136)., t_i is the treatment identifier (0= control; 1=intervention) and Q_i^b is the participant-specific baseline utility value (Manca, Hawkins and Sculpher 2005).

3-month cost-effectiveness analysis:

Table 25 summarises the output from the multiple regression analysis for the variable utility at month 3 adjusted by randomisation group and utility at baseline. Equation 4 summarises the trial-specific regression analysis in the estimate of the utility at month 3 adjusted by the randomisation group and utility at baseline.

Table 25. Summary of the regression analysis for utility at month 3 adjusted by two variables: randomisation group and baseline utility.

Utility month3	Coef.	Std. Err.	t	P>|t|	[95% Conf. Interval]	
Group	.0299768	.0382367	0.78	0.436	-.0465346	.1064882
Baseline utility	.4731861	.1159595	4.08	0.000	.2411516	.7052206
Constant	.4655077	.0991758	4.69	0.000	.2670573	.663958

Number of observations = 70; R-squared = 0.2581; Adjusted R-squared = 0.2329

Equation 4. Multiple regression analysis for the utility and month 3 adjusted per randomisation group and utility baseline.

*Utility at month 3 = 0.4655077 + 0.0299768 * Rand. Group + 0.4731861 * Utility at baseline*

Note: Rand. group (0 = control group; 1 = intervention group).

Based on Equation 4, for every unit increase in utility at baseline, a 0.473 increase in the utility at month 3 was estimated. Hence, based on the adjusted multiple regression, the utility difference between groups at month 3 is estimated at 0.0299768 (95% CI -0.0465346 to 0.1064882). Given the 3-month period considered, and assuming a linear interpolation, it is equivalent to a QALY differential of 0.0075 (95% CI -0.0116 to 0.0266). This compared with a differential utility of 0.0724 (95% CI -0.0102 to 0.15505) in the case where no adjustment for baseline utility was considered. The comparison between adjusted R-squared values (0.2329 vs 0.0328) for each model indicated that the adjusted model considering the utility at baseline better estimated the utility at month 3.

Cost-effectiveness was estimated based on incremental costs divided by incremental effects, in this case measured in QALYs. The mean cost per QALY at month 3 (Equation 5) was estimated at -£8,295. The intervention was dominant as it generated a higher number of QALYs at a lower cost.

Equation 5. Estimate of the incremental cost per QALY at month 3.

$$ICER = \frac{Cost\ intervention - Cost\ control}{QALY\ intervention - QALYcontrol} = \frac{£368.40 - £542.60}{0.216 - 0.195} = \frac{-£174.20}{0.021} = -£8{,}295$$

Figure 40 illustrates the bootstrap analysis with 1,000 replicates, based on the 3-month cost per QALY. At month 3, the intervention with MRI had a probability of 92.3% of being dominant (i.e. increased QALYs at a lower cost) compared to the control group. In the obverse quadrant, there was a 0.0% probability of the MRI intervention being dominated (i.e. lower QALYs at a higher cost) by standard care (i.e. control group). The remaining 7.7% of bootstraps were in the cost-effectiveness analysis quadrants, where the probability of being cost-effective also depends on the overall system willingness-to-pay for each QALY.

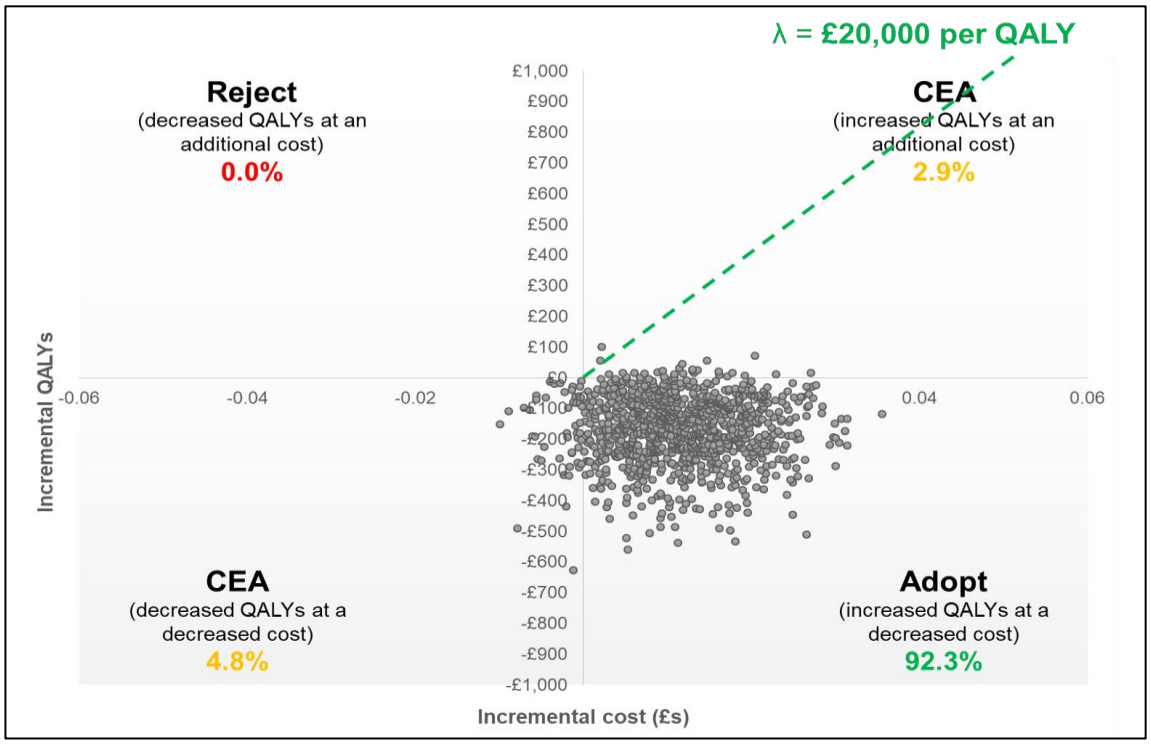


Figure 40. Cost-effectiveness plane associated with the 3-month cost per QALY analysis and probability associated each quadrant (bootstrap analysis with 1,000 replicates).

Using a £20,000 and £30,000 willingness-to-pay per QALY (thresholds typically considered by NICE) (McCabe, Claxton, and Culyer 2008), there was a 96.0% and 96.4% probability of MRI being cost-effective at month 3, respectively (Figure 41).

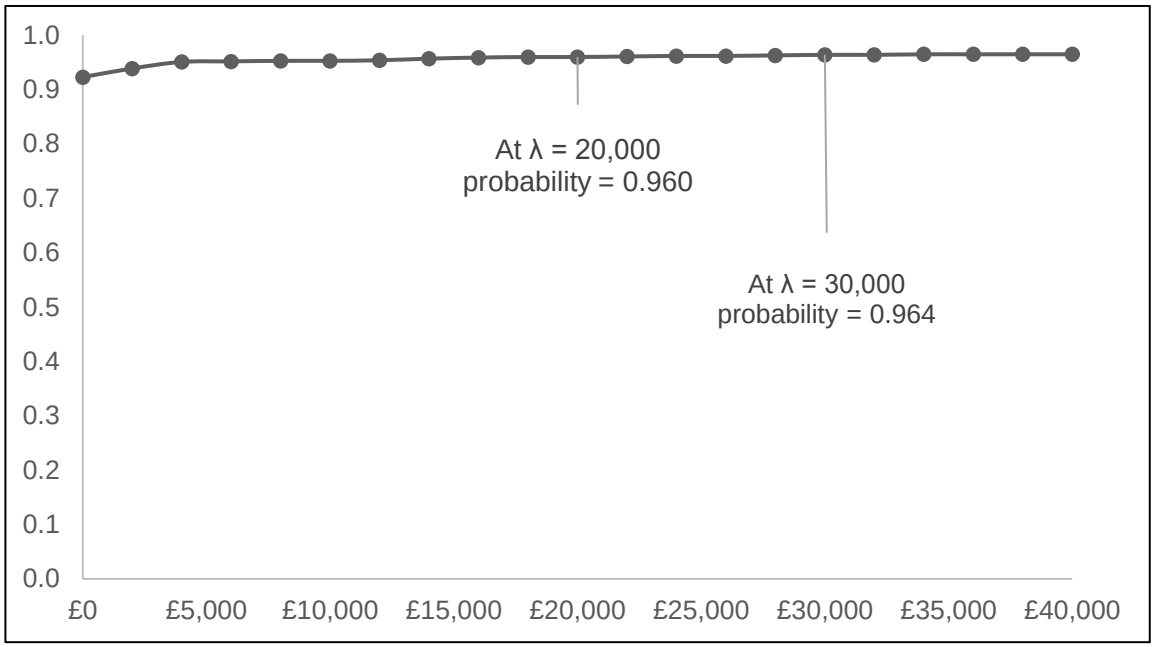


Figure 41. Cost-effectiveness acceptability curve for several thresholds of willingness-for-pay (represented as at λ).

EQ-5D-5L data completeness decreased during the follow-up period, with 60% of the data being complete at month 3 post-recruitment (Table 24). The utilisation of multiple imputation methods for missing utility data at months 1 and 3 did not affect the cost-utility analysis, with the intervention with MRI estimated at 95.5% and 96.2% probability of being cost-effective (£20,000 - £ 30,000 per QALY).

6-month cost-effectiveness analysis:

Table 26 summarises the output from the multiple regression analysis for the variable utility at month 6 adjusted by randomisation group and utility at baseline.

Table 26. Summary of the regression analysis for utility at month 6 adjusted by two variables: randomisation group; and utility baseline.

Utility month 6	Coef.	Std. Err.	t	P>|t|	[95% Conf. Interval]	
Group	.0278441	.02827	0.98	0.332	-.0298131	.0855012
Baseline utility	.0072316	.0687779	0.11	0.917	-.1330419	.1475051
Constant	.906974	.0609558	14.88	0.000	.7826539	1.031294

The mean cost per QALY at month 6 (Equation 6) was estimated at -£4,687. Again, the intervention was dominant as it generated a higher number of QALYs at a lower cost.

Equation 6. Estimate of the incremental cost per QALY at month 6.

$$ICER = \frac{Cost\ intervention - Cost\ control}{QALY\ intervention - QALY control} = \frac{£396 - £661}{0.452 - 0.395} = \frac{-£266}{0.057} = -£4{,}687$$

Figure 42 illustrates the bootstrap analysis with 1,000 replicates, based on the 6-month cost per QALY. At month 6, the intervention with MRI had a probability of 98.3% of being dominant and 0.0% of being dominated by the control group. The remaining 1.7% of bootstraps were in the cost-effectiveness analysis quadrants, i.e. the probability of being cost-effective also depends on the overall system willingness-to-pay for each QALY. Using a £20,000 and £30,000 willingness-to-pay per QALY (thresholds typically considered by NICE), there was a 100% probability of MRI being cost-effective at month 6 in both scenarios.

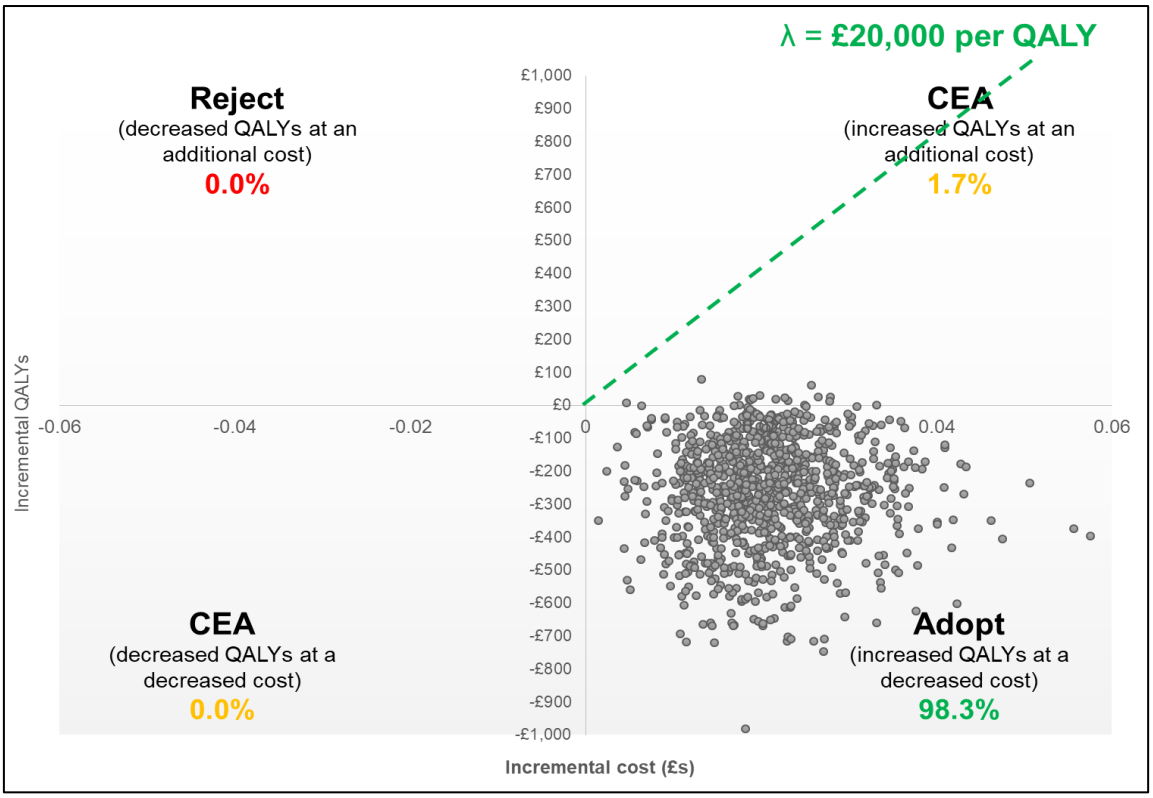


Figure 42. Probability associated with a participant being in a specific cost-effectiveness plane quadrant (1,000 bootstraps bootstrap analysis).

EQ-5D-5L data at 6 months were 58% complete (Table 24). The utilisation of multiple imputation methods for missing utility data at months 1, 3 and 6 did not affect the cost-utility analysis, with the intervention with MRI presenting a 99.95% and 99.97% probability of being cost-effective (£20,000 - £30,000 per QALY).

Pain score as the measure of effect:

A second cost-effectiveness analysis was based on self-reported pain scores as the measure of effect. The pain scores were based on a scale of 0-10, with 0 representing no pain at all and 10 the worst pain ever. The mean (SD) pain scores over a period of 3 months in the control and intervention group were, respectively, 3.61 (2.58) and 2.63 (1.59). The difference between groups approached statistical significance (p=0.074) (Table 27).

Table 27. Mean (SD) pain scores for participants in both randomisation groups.

Randomisation group	Mean pain score	Standard Deviation
Control group (n=22)	3.61	2.58
Intervention group (n=38)	2.63	1.59

The ICER using the incremental pain score as the measure of effect at month 3 indicated that the MRI intervention dominated the control group (ICER = £178 per pain score avoided), achieving better pain outcomes at a lower cost.

Outcome III. To estimate the mean cost per correctly diagnosed scaphoid fracture using MRI for patients with suspected scaphoid fracture and initial negative findings on conventional radiograph (with standard care as the comparator).

The mean cost per correctly diagnosed scaphoid fracture was estimated for both groups. This estimate included all diagnostic costs divided by the number of episodes where the correct diagnosis was achieved in the initial radiographs or MRI using the final 3-month radiographs as the reference. For instance, in the control group, if a scaphoid fracture was not seen in the follow-up period (including the final 3-month radiographs), then it was deemed that a correct diagnosis was reached on the initial x-ray. The same principle was applied to the intervention group.

The total and mean diagnostic costs for the control group were, respectively, £29,490 and £454. Out of 65 participants, a total of four scaphoid fractures were detected in the control group despite the initial negative x-ray. Hence, the correct diagnosis was achieved in 61 participants, leading to a mean cost per correct scaphoid fracture of £483.

The total and mean diagnostic costs for the intervention group were, respectively, £22,815 and £341. Out of 67 participants, no scaphoid fracture was diagnosed after the initial MRI exam. Hence, the mean cost per correct scaphoid fracture was estimated at £341. This is equivalent to a 30% reduction in the cost per correct scaphoid fracture compared to the control group (£341 vs £483).

If other fractures (e.g. radial or capitate bone) were also included, the number of correct diagnosed fractures in the control and intervention group were, respectively, 56 (5 non-scaphoid fractures diagnosed during the follow-up period) and 66 (1 non-scaphoid fracture – fracture of the hook of the hamate). Therefore, a mean cost per correctly diagnosed fractures were estimated at £526 and £346 for the control and intervention group, respectively. This is equivalent to a 34% reduction of the cost per correctly diagnosed bone fracture compared to the control group (£346 vs £526).

Outcome IV. To assess overall patient satisfaction associated with the proposed pathway in comparison to the current pathway.

Patient satisfaction was evaluated at month 3 post-recruitment and included three areas: (a) acute management of the suspected scaphoid fracture (immediately after recruitment) (Table 28); (b) elective management of the clinical condition (diagnostic and treatment pathway) (Table 29); and (c) participation in the research study (Table 30).

A higher proportion of participants in the intervention group were very satisfied with the ED episode compared to the control group (Table 28). Although both groups reported similar satisfaction levels regarding the overall ED visit (p=0.867), the intervention group was associated with a trend of better satisfaction levels regarding: how the injury was explained (question 1.1., p=0.088); the information received about the tests (question 1.2., p=0.075); and the information received about the test results (question 1.3, p=0.154). Following the discharge from the ED, participants in the intervention group reported trends of superior satisfaction levels concerning: the written information about their clinical condition (p=0.08).

With regards to the elective management of the suspected scaphoid fracture (Table 29), a higher proportion of participants in the intervention group compared to the control group were either very satisfied or satisfied (91.4% vs 80.0%) with the overall journey in the outpatient department. However, differences between groups were not statistically significant (p=0.482). The intervention was associated with improved trends of information received about any test result (p=0.114).

Participants were also asked about their experience in taking part in research (Table 30). There were no statistically significant differences between the groups when assessing different research components, particularly: information about the study explained to the participant (p=0.352); enough time to make an informed decision (p=0.436); clear understanding of the aims (p=0.281); enough opportunities to ask questions (p=0.854); better understanding of the clinical condition due to taking part in the study (p=0.333). Participants in the intervention group showed a statistically significant improved satisfaction from taking part in the trial (p=0.043). Likewise, participants in the intervention group reported that they believed that taking part in the study improved their clinical care (81% vs 19%, p<0.001).

Table 28. Patient experience questionnaire for the acute management of the pathway in the control (n=22) and intervention (n=41) groups.

		Very satisfied		Satisfied		Neutral		Dissatisfied		Very dissatisfied	
		N	%	N	%	N	%	N	%	N	%
1. Presentation to the Emergency Department (ED)											
1.1. How well your injury was explained to you by staff	Control group	9	43%	10	48%	1	4.8%	1	4.8%	0	0.0%
	Interv. group	31	76%	8	20%	1	2.4%	1.0	2.4%	0	0.0%
1.2. The information you received about any tests you needed	Control group	7	33%	12	57%	1	4.8%	1	4.8%	0	0.0%
	Interv. group	28	68%	11	27%	1	2.4%	1.0	2.4%	0	0.0%
1.3. The information you received about the results of any tests	Control group	9	47%	7	37%	2	11%	1	5.3%	0	0.0%
	Interv. group	23	56%	12	29%	1	2.4%	5	12%	0	0.0%
1.4. How you found the visit overall	Control group	11	52%	8	38%	1	4.8%	1	4.8%	0	0.0%
	Interv. group	22	54%	14	34%	4.0	9.8%	1	2.4%	0	0.0%

		Yes		No, but I would have liked it		No, but I do not need this type of information	
		N	%	N	%	N	%
1.5. Were you given any printed information about your condition to take away with you?	Control group	13	62%	6	28%	2	9.5%
	Interv. group	33	81%	3	7.3%	5	12.2%
1.6. Did a member of staff tell you when you could resume your usual activities, such as when to go back to work or drive a car?	Control group	11	52%	5	24%	5	24%
	Interv. group	30	73%	4	9.8%	7.0	17%

		Yes		No		Don't know	
		N	%	N	%	N	%
1.7. Did hospital staff check if you would be adequately supported at home when you were leaving the ED?	Control group	12	57%	4	19%	5	24%
	Interv. group	26	65%	9	23%	5	13%
1.8. Did hospital staff tell you who to contact if you were worried about your condition or treatment after you left the ED?	Control group	14	67%	5	24%	2	9.5%
	Interv. group	31	78%	4	10%	5	13%

Table 29. Patient experience questionnaire for the elective component of the pathway in the control (n=22) and treatment (n=41) groups.

		1 to 2 weeks		2 weeks to 1 month		1 to 2 months		2 to 3 months	
		N	%	N	%	N	%	N	%
2. Outpatient follow-up care									
2.1. Following on from your visit to the ED, when did your first outpatient appointment take place?	Control group	12	57%	5	24%	2	9.5%	2	9.5%
	Interv. group	16	42%	1	2.6%	5	13%	16	42%

		1 visit		2 visits		3 visits	
		N	%	N	%	N	%
2.2 How many outpatient visits did you make in total between your first visit to an ED and your final x-ray at 3 months?	Control group	7	33%	3	14%	11	52%
	Interv. group	13	42%	8	26%	10	32%

		Very satisfied		Satisfied		Neutral		Dissatisfied		Very dissatisfied	
		N	%	N	%	N	%	N	%	N	%
2.3 Thinking about your visit(s) to the Outpatient Department in general, how did you find the following aspects of your care?											
The information you were given to manage your injury at home	Control group	5	25%	10	50%	3	15%	1	5.0%	1	5.0%
	Interv. group	18	50%	14	39%	3	8.3%	1	2.8%	0	0.0%
The information you received about any extra tests you needed	Control group	4	20%	11	55%	5	25%	0	0.0%	0	0.0%
	Interv. group	13	39%	14	42%	5	15%	1	3.0%	0	0.0%
The information you received about the results of any tests	Control group	3	17%	10	56%	3	17%	2	11%	0	0.0%
	Interv. group	17	52%	11	33%	3	9.1%	2.0	6.1%	0	0.0%
How you found your visits to Outpatients overall	Control group	8	40%	8	40%	3	15%	1.0	5.0%	0	0.0%
	Interv. group	17	49%	15	43%	3	8.6%	0.0	0.0%	0	0.0%

Table 30. Patient experience questionnaire for taking part in the trial for participants in the control (n=22) and treatment (n=41) groups.

		Strongly agree		Agree		Neutral		Disagree		Strongly disagree	
		N	%	N	%	N	%	N	%	N	%
3. Participating in the study											
3.1. How far would you agree with the following statements on your experience of taking part in this study?											
I was satisfied with the way the information about the study was explained to me	Control group	11	52%	8	38%	2	9.5%	0	0.0%	0	0.0%
	Interv. group	29	71%	10	24%	2	4.9%	0	0.0%	0	0.0%
I had enough time to make an informed decision to participate in the study	Control group	11	52%	7	33%	3	14%	0	0.0%	0	0.0%
	Interv. group	24	59%	15	37%	2	4.9%	0	0.0%	0	0.0%
I have a clear understanding of the study aims	Control group	10	48%	9	43%	2	9.5%	0	0.0%	0	0.0%
	Interv. group	28	68%	11	27%	2	4.9%	0	0.0%	0	0.0%
I had enough opportunities to ask questions and for support if needed over the duration of the study	Control group	9	43%	8	43%	2	9.5%	1	4.8%	0	0.0%
	Interv. group	22	54%	15	37%	3	7.3%	1	2.4%	0	0.0%
I had a good overall experience of taking part in this study	Control group	9	43%	8	38%	4	19%	0	0.0%	0	0.0%
	Interv. group	27	68%	12	30%	1	2.5%	0	0.0%	0	0.0%

		It improved my care		It made no difference		It had a negative impact	
		N	%	N	%	N	%
3.2. Do you think taking part in the study had any impact on your care?	Control group	4	19%	16	76%	1	4.8%
	Interv. group	33	81%	8	20%	0	0.0%

		It improved my understanding		It made no difference		It had a negative impact	
		N	%	N	%	N	%
3.3. Do you feel that taking part in the study changed your understanding of your condition and treatment?	Control group	14	67%	6	29%	1	4.8%
	Interv. group	31	76%	10	24%	0	0.0%

Outcome V. To estimate the accuracy of the proposed intervention (i.e. with scaphoid MRI) in the detection of scaphoid fracture compared to the current pathway (i.e. 4-view radiographs only).

The accuracy of the intervention (immediate acute wrist MRI) and the standard care were compared against the 3-month radiographs (assumed to be the reference). Accuracy was estimated by calculating the number of correct diagnoses (either true positives – "there is a fracture" – or true negatives – "there is not a fracture") divided by the total number of assessments, in this case the number of participants (as summarised in Equation 7).

Equation 7. General accuracy equation and its respective estimate for both control and intervention (MRI) groups in the detection of scaphoid fractures.

$$Accuracy = \frac{number\ of\ True\ Negatives + number\ of\ True\ Positives}{total\ of\ asessments}$$

$$Accuracy\ in\ the\ diagnosis\ of\ scaphoid\ fractures\ at\ ED\ (control\ group) = \frac{61 + 0}{65} = 93.8\%;$$

$$Accuracy\ in\ the\ diagnosis\ of\ scaphoid\ fractures\ at\ ED\ (MRI\ group) = \frac{60 + 7}{67} = 100\%;$$

As expressed in Equation 7, the accuracy for the detection of scaphoid fractures in the control group, which is determined by the initial radiographs only, was estimated at 93.8%. This contrasts with the 100% figure in the intervention group, i.e. the MRI did not generate any false negatives (participants with scaphoid fracture that tested negative) or false positives (participants with no scaphoid fractures that tested positive).

When other fractures (e.g. radial or capitate bone) are also included, the accuracy for detection for any fractures decreased to 84.6% and 98.5% in the control and intervention groups, respectively (see Equation 8).

Equation 8. General accuracy equation and its respective estimate for both control and intervention (MRI) groups in the detection of any bone fracture.

$$Accuracy\ in\ the\ diagnosis\ of\ any\ bone\ fracture\ at\ ED\ (control\ group) = \frac{55 + 0}{65} = 84.6\%;$$

$$Accuracy\ in\ the\ diagnosis\ of\ any\ bone\ fracture\ at\ ED(MRI\ group) = \frac{47 + 19}{67} = 98.5\%;$$

The intervention group had one false negative finding, as one participant with a normal wrist MRI was subsequently found to have a fracture of the hook of the hamate. Retrospectively, unaware of the original report, a second musculoskeletal radiologist was asked to report the MRI. The second radiologist reported bone oedema in the hook of hamate, consistent with a potential fracture seen in the initial wrist MRI sagittal plane (Figure 43a) as well as the subsequent wrist CT scan (Figure 43b). This suggests that the false negative may have been associated with the reporting rather than the MRI imaging dataset.

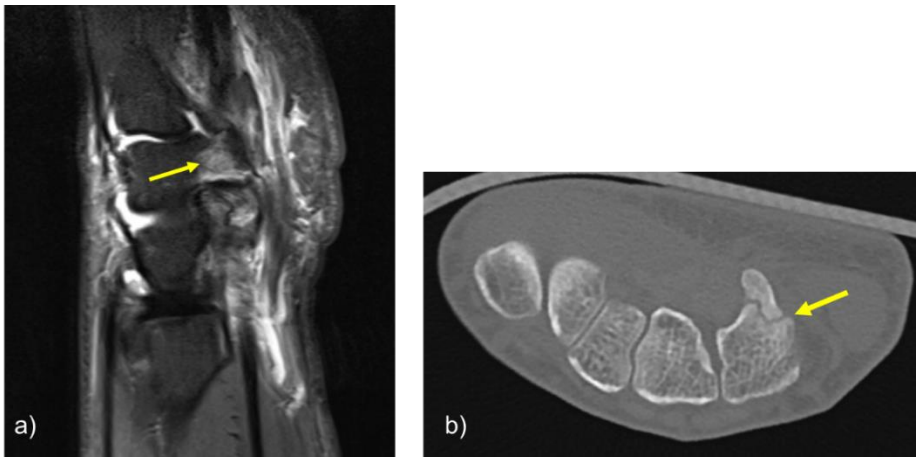


Figure 43. (a) Bone oedema seen in the hook of the hamate on original sagittal MRI; and (b) hook of hamate fracture demonstrated in wrist CT.

Outcome VI. To estimate the time taken to reach a definitive diagnosis in the intervention group compared to the control group.

The time taken to reach a definitive diagnosis was considered to be once a scaphoid fracture was effectively ruled-out or ruled-in. For instance, if immediate MRI correctly ruled-out a scaphoid fracture on presentation to the ED, then that was the time a definitive diagnosis was reached. In contrast, if the findings in the initial MRI were found to be inaccurate (e.g. subsequent imaging found a scaphoid fracture), the latter time was taken to be the definitive diagnosis.

The time taken to reach a definitive diagnosis (measured in days) was estimated at 10.2 (0 to 55) and 1.7 (0 to 116) in the control and intervention group, respectively. The MRI intervention led to a quicker definitive diagnosis ($p<0.001$).

The time spent in the ED was also evaluated. The dataset was almost complete with only data for one participant in each group missing as both participants left the ED prior to formal discharge. The mean time elapsed in the ED (measured in hours: minutes) for participants in the control and intervention group were, respectively, 2:12 (1:03) and 3:20 (1:01). The MRI intervention led to participants staying on mean 68 minutes longer in the ED ($p<0.001$).

Outcome VII. To estimate the amount of time off work or informal care needs due to the suspected scaphoid fracture.

The time participants spent immobilised with a plaster cast was measured as this variable is commonly used as a proxy to assess the potential societal impact of the intervention. A total of 14 and 20 plaster casts were used in the control and intervention groups, respectively. The latter was due to the higher proportion of fractures (scaphoid or otherwise) and significant soft issue injuries detected in the intervention group. The mean number of days immobilised (SD) with plaster cast among patients that needed a plaster cast were 25.9 (8.2) and 36.1 (11.6) days in the control and intervention groups, respectively, but the difference was not statistically significant ($p=0.397$).

Time off work or informal care due to the suspected scaphoid fracture episode were also assessed. One hypothesis behind the intervention was that the use of immediate MRI may be associated with a lower proportion of participants being immobilised and thus requiring time off work or informal care. The mean (SD) time off work and informal care in the control (n=22) and the intervention groups (n=38) were, respectively, 6.0 (9.42) vs 4.3 (9.55) days. This difference was not statistically significant (Mann-Whitney U test, p=0.249).

3.3.7 Sensitivity Analyses

Deterministic sensitivity analyses around three model unit cost parameters were performed: (i) immediate wrist MRI in the ED; (ii) fracture clinic appointments (both first and follow-up outpatient appointments); and (iii) 'Did not attend' cost. Two additional scenarios were considered: (iv) using existing reimbursement strategies as a proxy of unit costs; and (v) adopting a societal perspective of analysis instead of the NHS and Personal and Social Services perspective. Table 31 and Table 32 summarise the impact in the 3 and 6-month cost analyses and 6-month cost-effectiveness, respectively.

(i) Immediate wrist MRI in the ED. The unit cost per immediate wrist MRI in the ED was a fundamental variable associated with the provision of the intervention. The base case scenario used a unit cost of £72.40, based on: (a) the unit cost of the non-acute wrist MRI estimated at £120.73 (estimate retrieved from NHS Reference Costs 2017-18 code RD01A - Magnetic Resonance Imaging Scan of One Area, without Contrast, 19 years and over); and (b) the time it took to complete the acute short-sequence wrist MRI (approximately 15 minutes) compared to the conventional wrist MRI scan (approximately 25 minutes). Hence, the acute short-sequence wrist MRI was estimated as a proportion of the conventional wrist MRI. This assumption applied not only to the time it took to complete the actual acquisition time but also the subsequent reporting.

This deterministic sensitivity analysis assumed that the unit cost of the intervention (abbreviated wrist MRI) would equal the cost of the full wrist MRI (£120.73). Under this scenario, the cost difference per participant at 3 and 6 months between groups decreased to -£129 (p=0.221) and -£220 (p=0.105). At 6 months, the intervention remained dominant and cost-effective at both willingness-to-pay thresholds (£20,000 and £30,000 per QALY).

(ii) Fracture clinic appointments. The unit cost per fracture clinic appointment (first and follow-up appointments) was varied by -25%. The rationale behind this sensitivity analysis was to estimate the potential downstream impact of using MRI as an add-on initial test. A -25% unit cost variation led to a decrease of the cost difference per group, estimated at -£134 (p=0.157) and -£234 (p=0.073) per participant at 3 and 6 months, respectively. In both scenarios, the MRI intervention remained dominant and cost-effective at both willingness-to-pay thresholds.

(iii) 'Did not attend' cost. The unit cost per 'Did not attend' (DNA) event was assumed at 50% of the base case cost of the event when patients did attend and was particularly relevant in outpatient visits (i.e. Fracture Clinic appointments). A -25% variation led to a mean cost difference between groups of -£166 (p=0.112) at month 3 and -£258 (p=0.055) at month 6. In both scenarios, the MRI intervention remained dominant and cost-effective at both willingness-to-pay thresholds.

(iv) Existing reimbursement strategies. The use of existing reimbursement strategies as a proxy for measuring the costs from a NHS and Personal Social perspective was also explored. The latter means that instead of valuing individual costs associated with each event, the reimbursement schedules used to pay healthcare providers for these healthcare events were used instead. This sensitivity analysis is relevant as existing NHS reimbursement and incentive strategies might not be aligned with the proposed intervention. Ultimately, the intervention might generate cost savings to the NHS but simultaneously generate a loss to individual healthcare providers, thus compromising the implementation of the intervention and its inclusion in routine clinical practice. As one example, the use of advanced imaging in the Fracture Clinic is considered to be part of a bundled payment, i.e. the use of CT or MRI in Fracture Clinic is included in the Fracture Clinic tariff and therefore there is no increase in income received by the healthcare provider, in this case GSTT.

This sensitivity scenario led to a mean cost difference between groups of -£185 (p=0.166) at month 3 and -£273 (p=0.184) at month 6. As with the other sensitivity scenarios, the MRI intervention remained dominant and cost-effective at both willingness-to-pay thresholds.

(v) Societal costs. The broadening of the perspective of analysis was also considered to the societal perspective rather than a NHS and PSS perspective (base case). This analysis included all non-NHS costs, such as healthcare costs that happened within the private sector and any costs associated with time off work. From a societal perspective, the mean cost per participant (SD) between groups increased to -£754 (p=0.101), with the intervention remaining dominant and cost-effective at both willingness-to-pay thresholds.

Table 31. Sensitivity analyses scenarios and impact in the cost analyses at month 3 and 6.

	Mean cost difference		**p-value (using GLM)**	
	Month 3	Month 6	Month 3	Month 6
Base case scenario	**-£174**	**-£266**	**0.094**	**0.047**
(i) **Immediate wrist MRI in the ED:** equals to full wrist MRI unit cost (£120.74)	-£129	-£220	0.221	0.105
(ii) **Fracture clinic appointments:** **-25%** of unit cost of the initial/follow-up Fracture Clinic	-£143	-£234	0.157	0.030
(iii) **'Did not attend' cost** **-25%** of the absolute proportion of DNA cost compared to event where the patient attended (baseline=50%).	-£166	-£258	0.112	0.055
(iv) **Existing reimbursement strategies.** The utilisation of **existing reimbursement strategies** to pay for clinical care provided	-£185	-£273	0.166	0.184
(v) **Societal costs.** The use of a **societal perspective of analysis**.	-£754		0.101	

Table 32. Sensitivity analyses scenarios and respective impact on cost-effectiveness at month 6 and probability of being cost-effective at NICE willingness-to-pay thresholds.

	Control group vs intervention group at month 6				
	ICER (£/QALY)			**Probability of cost-effectiveness at given threshold**	
	Mean difference in costs (£)	**Mean difference in QALYs**	**ICER**	**£20k**	**£30k**
Base Case	-£266	0.057	Intervention Dominant	100%	100%
(i) Immediate wrist MRI in the ED: equals to full wrist MRI unit cost (£120.74)	-£220	0.057	Intervention Dominant	100%	100%
(ii) **Fracture clinic appointments**. The unit cost per fracture clinic appointment (first and follow-up appointments) was decreased by 25%	-£234	0.057	Intervention Dominant	100%	100%
(iii) **'Did not attend' (DNA) cost**. Proportion of DNA cost reduced to 25% of the overall unit cost of the episode (base case of 50%)	-£258	0.057	Intervention Dominant	100%	100%
(iv) **Existing reimbursement strategies.** The utilisation of **existing reimbursement strategies** to pay for clinical care provided	-£273	0.057	Intervention Dominant	100%	100%
(v) **Societal perspective** of analysis.	-£754	0.057	Intervention Dominant	100%	100%

3.3.8 Summary of Results

The MRI intervention led to lower mean costs per participant at both 3 and 6 months post-recruitment. Using GLM distributions, the mean cost difference per participant between both groups at 3 and 6-months post-recruitment were estimated at -£174 (CI 95%: - £378 to £30) and -£266 (CI 95%: -£528 to -£3), respectively. Although the use of immediate MRI led to cost savings, these were only statistically significant at 6 months. However, if bias-corrected bootstrap analyses are considered, the cost difference was also significant at month 3 (95% CI: -£487; -£20).

Assuming £20,000 and £30,000 willingness-to-pay per QALY thresholds, there was a 96.0% and 96.4% probability of MRI being cost-effective at month 3, compared to a 100.0% probability at month 6.

A higher number of major injuries (scaphoid or other bone fractures or soft tissue injuries) were detected in the initial MRI, allowing for more targeted treatment of patients that required wrist immobilisation. The MRI intervention was associated with higher accuracy levels, leading to a better and quicker diagnosis, and some areas of improvement in patient satisfaction. However, the intervention also led to an increase in the mean time elapsed in the ED as part of the initial presentation.

3.4 Discussion

3.4.1 Overview

This section discusses the clinical and economic findings from the SMaRT trial. Furthermore, strengths and limitations are also debated, alongside potential implications of this trial both in the field of research and clinical practice associated with the management of this clinical condition.

3.4.2 Findings

Study design

To our knowledge, this was the first randomised trial that evaluated the clinical and cost implications of using immediate MRI in the acute management of suspected scaphoid fractures with negative findings on the initial radiographs. This feature was innovative, as previous empirical studies had only evaluated the use of advanced imaging two to five days after presentation to the ED (Rua et al. 2017). Given its innovative nature, there was some uncertainty regarding the operational feasibility of providing MRI in an acute setting, i.e. as part of an ED pathway. To address this, and prior to the SMaRT trial, a 20-patient pilot was conducted to assess the trial's feasibility and inform about its design, particularly feedback regarding the trial's documentation and workflows.

Participant baseline characteristics

Several participant characteristics were captured at baseline, prior to any randomisation related process. This approach was necessary to avoid any potential bias in self-reported data captured

by the EQ-5D-5L questionnaire. As specified in the *a priori* statistical analysis plan, no significance testing on the baseline variables between intervention groups was used given the trial's randomised design. Furthermore, although the variable number of wrists was initially considered as a stratification variable, this was not deemed feasible due to the very low number of participants with suspected injuries to both wrists (n=3).

The mean (SD) age per participant in the control and intervention groups were 36.2 (12.6) and 38.2 years (13.4), respectively. This young age is consistent with the clinical literature, as the incidence of suspected clinical fractures is commonly associated with sports-related injuries (Yin et al. 2010). With regards to gender, the number (%) of male participants were 34 (52%) and 41 (61%) in the control and intervention groups, respectively.

In relation to employment status, a total number (%) of 51 (78%) and 53 (79%) participants were employed full-time (30 hours or more) in the control and intervention groups, respectively. This situation has implicit economic costs to society, evaluated as part of the sensitivity analysis.

Clinical findings

A higher number of scaphoid fractures were detected in the intervention group (7 fractures, incidence of 10.4%) compared to the control group (4 fractures, incidence of 6.2%). However, this difference was not found to be statistical significant (p=0.37). These findings were consistent with the systematic review performed by Yin et al. (2010). All seven scaphoid fractures in the intervention group were found at baseline (day 0) whilst in the control group the diagnosis was only made, on mean, 20 days (SD=19.1) following the presentation to ED. One participant in each group required surgical repair associated with the scaphoid fracture.

Similarly, the intervention group also reported a higher number of fractures in other bone (excluding the scaphoid bone), with 15 (incidence of 22.4%) bone fractures compared to 5 (incidence of 7.7%) in the control group. Unlike the detection of scaphoid fractures, this difference was statistically significant (p=0.019). Out of the 22 fractures detected in other bones, 50% were radial fractures, with 4 and 7 in the control and intervention groups, respectively.

Additionally, the MRI intervention allowed the diagnosis of soft tissue injuries that otherwise would not be detected during the initial ED episode. Four soft tissue injuries were detected, with emphasis on ligament tears in the triangular fibrocartilage complex (TFFC). However, none of these injuries required surgical repair as opposed to the only significant soft tissue detected in the control group. This injury, a complete scapholunate ligament rupture, was detected on CT eleven days after the ED presentation and required two surgical repairs (initial procedure to perform reconstruction procedures and the second for the removal of k-wires).

Given the trial's randomised design, a similar incidence of scaphoid or other bone fractures was anticipated. However, that is not the case as the intervention with MRI was associated with a higher proportion of injuries being diagnosed. This situation might be due to three clinical scenarios: (i) injuries not being diagnosed in the control group; (ii) non-relevant injuries being diagnosed in the intervention group; or (iii) a mixture of both phenomena. First, clinically relevant injuries such as

soft tissue injuries might remain undiagnosed in the control group, especially if participants are not clinically symptomatic. As an example, one participant randomised to the control group was symptomatic at month 3 with a TFFC injury. Second, even minor lesions were diagnosed due to the increased accuracy levels of the immediate use of MRI during the initial ED episode. Some of the fractures detected in the intervention group were trabecular scaphoid fractures that did not require immobilisation with plaster cast and would not have been diagnosed in the control group unless advanced imaging (CT or MRI) was performed. Given that the control group, i.e. standard care, included the splint immobilisation of all participants, it can be assumed that some injuries diagnosed in the intervention group had minimal clinical impact. Nevertheless, diagnosis of such lesion allowed a targeted follow-up diagnostic pathway in order to ensure the appropriateness of the healing process.

As a corollary, the use of immediate MRI led to a quicker diagnosis, allowing the detection of a higher number of bone fractures (scaphoid or otherwise) and soft tissue injuries.

Primary outcome:

3-month cost analysis

The primary outcome was total costs at 3 months post-recruitment associated with the proposed intervention (i.e. immediate wrist MRI as an add-on test) compared to standard care that relied on the radiographs only during the ED episode. These costs were estimated taking a NHS and Personal and Social Services perspective. The choice of primary outcome was due to the absence of economic evidence associated with the proposed intervention. The 3-month timeline was deemed appropriate to capture all relevant NHS resource use given the short-term nature of the clinical condition. As with most cost datasets, these data were expected to be positively skewed. Hence, the *a priori* statistical plan used GLM and bootstrap analysis for the primary and secondary statistical analyses.

The trial found a trend for the intervention to be associated with lower costs per participant (-£174), leading to total savings of £11,832 compared to the control group. However, the significance of this depended on the statistical method employed. Both methods are recommended to assess skewed data such as costs (Glick et al. 2007; Gray et al. 2011) and were specified in the *a priori* statistical analysis plan to estimate the mean cost difference between groups, with GLM being the primary method and bootstrap the secondary. The primary GLM analysis found the cost difference to be non-statistically significant (95% CI: -£378, £30, p=0.094) whilst with the 1000-replicate bias-corrected bootstrap analysis the cost difference was statistically significant (95% CI: -£486, -£20). Given its potential impact to affect the intervention's adoption decision, further research should be considered to determine the more adequate statistical method to evaluate cost datasets.

The superior clinical accuracy of MRI compared to conventional radiographs had two contrasting cost implications. On one hand, due to its superior sensitivity, MRI detected a higher number of clinical conditions, some of these warranting treatment, and hence leading to an increase in costs. Due to the MRI's superior specificity, immediate MRI was able to rule-out major injuries (bone

fractures and soft tissue injuries) thus reducing the need for follow-up and decreasing the respective costs. In the end, due to the low overall incidence of major injuries, the MRI's ability to quickly rule-out injuries prevailed, leading to a decreased cost per participant in the MRI group.

The intervention (MRI) group also had a higher number (%) of participants that fell in the low cost £0-£250 range, with 27 (40%) and 11 (17%) participants in the MRI and control group, respectively. However, the minimum cost in the control group (£94, n=4) was lower than in the intervention group (£166, n=22). These four participants in the control group had no formal follow-up as either cancelled the appointment (n=2) or left the ED prior to having one appointment booked (n=2). The 22 participants in the intervention group who had any major injury ruled-out in the ED and had no subsequent follow-up whatsoever. This was indeed the rationale for the intervention as the upfront costs of providing immediate MRI avoided downstream costs associated with outpatient appointments and further diagnostic tests. The intervention and control groups had similar number (%) of participants that cost between £251 and £500 with 26 (40%) and 25 (37%) participants, respectively. These are typically participants that required one initial fracture clinic appointment and none or one follow-up appointment. The cost distribution in the control group is more positively skewed than the intervention group, with 28 (43%) participants in the control group with costs between £500-£1000 compared to only 13 (19%) in the intervention group. The latter is due to the higher number of appointments performed to either rule-in or rule-out a scaphoid fracture in the control group. Finally both groups had only one participant that incurred in costs over £1000. These two participants required surgical repair of a complete scapholunate ligament injury (control group) and an undisplaced fracture of the body of the scaphoid (intervention group). However, the first surgical repair entailed specialised reconstruction procedures divided in two actual surgical procedures, resulting in a maximum cost of £7,116 and £2,691 in the control and intervention group, respectively.

The intervention with MRI was associated with a lower number of NHS events compared to the control group, both in primary and secondary care. In terms of primary care utilisation, the intervention led to a statistically significant lower number of GP face-to-face appointments. With regards to secondary care, compared to the control group, the intervention led to a reduced utilisation of both initial and follow-up fracture clinic appointments as well reduced use of advanced imaging (CT or MRI). In fact, almost half of participants in the control group (30/65) ended up undergoing advanced imaging, either CT or MRI. The latter meant that, more than just increasing the utilisation of advanced imaging, the immediate use of MRI shifted the use of these technologies to the beginning rather than the end of the diagnostic pathway. Simultaneously, the use of immediate wrist MRI led to the release of slots from GPs and orthopaedists, time that can be used to manage other patients that would actually require care.

Secondary outcomes

6-month cost analysis. Mindful of the potential limitations associated with the 3-month timeline, one secondary outcome considered the extension of the cost analysis up to 6 months post-randomisation. The mean cost difference between the two groups at month 6 was higher compared

to month 3 (£266 vs £174). In addition, the cost difference was found to be statistically significant using both the GLM analysis (95% CI: -£528, -£3, p=0.047) and the 1000-replicate bias-corrected bootstrap analysis (95% CI: -£635, -£29). The difference from the two groups derived mainly from just one participant in the control group that had surgery after the initial 3 months. Although the correct diagnosis of the injury was achieved 48 days post-randomisation, the surgical repair was delayed on multiple occasions due to an ongoing multi resistant infection (MRSA). Hence, the high cost of this participant (£6,986) was not included in the primary outcome (at month 3) but was in the secondary outcome (at month 6). This difference in resource utilisation at follow-up, particularly in terms of fracture clinic appointments (initial and follow-up), advanced imaging (MRI and CT) and surgeries drove the cost difference and allowed the recoup of the costs associated with the intervention. Given the low number and high unit cost of surgeries, further cost implications were assessed. If only NHS costs incurred up to the moment of definitive diagnosis were included, thereby excluding any treatment costs, particularly surgical procedures costs, the cost difference between the two groups remained statistically significant, with the intervention leading to mean cost savings per participant of £113 (CI 95%: - £188 to - £39). This analysis supported the findings supporting the cost savings to the NHS generated by the innovative use of advanced imaging.

In summary, at 6 months post-recruitment and regardless of the statistical method used, the MRI intervention was associated with lower costs per participant. These cost findings from the SMaRT trial diverged from existing randomised clinical trials, given that a statistical significant cost difference between groups was found (Kelson, Davidson, and Baker 2016; Patel et al. 2013; Brooks et al. 2005). This may be due to the fact that: (i) the SMaRT trial was powered to capture cost differences; and (ii) the use of immediate MRI was used to streamline the clinical pathway and reduce secondary care appointments among patients without any bone fracture.

3 and 6-month cost-effectiveness. In addition to the cost analyses described, cost-effectiveness analyses at month 3 and 6 regarding the intervention were also conducted using two measures of effect: (i) QALYs; and (ii) correct diagnosis per suspected scaphoid fracture. No significant differences in utilities between groups were detected at baseline. This was consistent with the trial's randomised design. Over the follow-up period, participants in the intervention group reported higher quality of life compared to the control group (a trend at month 3). However, a proportion of follow-up utility data were missing, with a maximum of 42% missing data at month 6. Given its potential impact, particularly in the cost-utility analyses, we evaluated whether the follow-up data were missing at random. Although no statistical differences were detected, the presence or absence of follow-up data seemed to affect the 6-month costs (p=0.131), with participants with no follow-up data reporting higher cost differences. For this reason, multiple imputation methods were used to deal with the missing utility data.

The MRI intervention was highly likely to be cost-effective at the traditional willingness-to-pay thresholds considered by NICE (£20,000-£30,000). This result was mainly driven by the cost reduction associated with the intervention, with a probability of 96% and 100% of being cost-

effective at month 3 and 6, respectively. Moreover, the utilisation of multiple imputation methods for utility data missing at month 1, 3 and 6 did not affect these cost-utility results.

The second cost-effectiveness analysis considered the cost per correctly diagnosed suspected scaphoid fracture. As with the incremental cost per QALY, the intervention is also associated with a lower mean cost per correct diagnosis (£483 vs £341), equivalent to a 30% reduction. Hence, irrespective of the measure of effect considered, immediate MRI was cost-effective. Despite the immediate nature of the intervention (on the day of presentation to the ED) compared to 2-5 days after presentation from previous evidence (Yin, Zhang, and Gong 2015; Karl, Swart, and Strauch 2015), this finding clarified the role of advanced imaging in the management of scaphoid fractures as evidence from previous cost-effectiveness and cost-utility studies was inconclusive and presented multiple serious methodological issues. Some of these methodological issues comprised lack of empirical data, proposed clinical pathways not consistent with the real-world deployment of advanced imaging, and other significant clinical findings such as other bone fractures or soft tissue injuries not being considered.

Diagnostic accuracy. The accuracy of radiographs and immediate MRI at diagnosing scaphoid fractures in the ED were 93.8% and 100%, respectively. This meant that MRI was able to correctly either rule-out or rule-in scaphoid fractures in all patients presenting with a clinical suspicion of scaphoid fracture and negative findings on the initial conventional radiograph. In contrast, in 6.2% of the cases, radiographs reported false negative findings, i.e. no findings reported on the imaging scan despite the patient's actual scaphoid fracture. If we extend this estimate to all bone fractures, conventional radiographs only and MRI have reported accuracies of 84.6% and 98.5%, respectively. Hence, immediate MRI contributed to a significant improvement associated with the accuracy of the diagnostic component of the pathway, leading to only one false negative result compared to ten in the control group.

Time from ED admission to discharge. The time elapsed in the ED was estimated as providing immediate MRI in an acute setting - constrained with provision of care within specific time targets - was thought to be challenging. Moreover, this variable was deemed relevant as not meeting existing 4-hour ED targets can lead to financial penalties to the healthcare provider.

The median time (measured in hours : minutes) from ED presentation to discharge was 02:10 and 03:17 for the control and intervention group, respectively. The mean time from ED presentation to discharge was also similar, with 02:12 and 03:21 for the control and intervention groups, respectively. A total of 3 (3.1%) and 14 (20.6%) ED episodes took more than 4 hours from ED presentation to discharge. The implementation of immediate MRI was associated with an increased median and mean ED time of 01:07 and 01:08 ($p<0.001$), respectively. This time difference between groups was consistent with the time from randomisation to ED discharge. Hence, the time difference appeared to be directly attributable to the intervention, i.e. the processes associated with the provision of the MRI scan and its subsequent reporting. The potential implementation of this clinical pathway needs to take into consideration the potential breaches of the NHS 4-hour ED target and how to deal with this process outcome. Given that this might be associated with negative

financial incentives for the healthcare provider, real-world implementation of the intervention needs to be discussed with regulatory agencies in order to incentivise the use of a technology with overall improved clinical and financial outcomes (see Chapter 7 for further detail on the implementation plan). It is relevant to highlight that these potential negative incentives, as well as scaphoid-related litigation costs (scaphoid-related injuries are one the main reasons for litigation costs at GSTT), were not considered in the cost analyses.

Time taken to reach a diagnosis. The time taken to reach a definitive diagnosis in both groups was also estimated. The rationale is that the use of MRI would lead to a quicker definitive diagnosis and, if needed, treatment. The mean time (SD) taken to reach a definitive diagnosis in the control and intervention groups were, respectively, 10.5 (6.5) and 1.7 (14.2) days ($p<0.001$). The latter is due to MRI's higher ability to either rule-in or rule-out a bone fracture in participants presenting with a suspicion of scaphoid fracture.

Patient satisfaction. The evaluation of self-reported patient satisfaction was considered as the intervention holds the potential to impact different dimensions of analysis. Regardless of the randomisation group, participants were satisfied by the acute and elective management of the suspected scaphoid fracture. Nonetheless, participants in the intervention group exhibited a trend of higher satisfaction levels compared to the control group. Despite being for longer in the ED and submitted to further tests, participants in the intervention group reported similar levels of satisfaction compared to the control group. With regards to the elective follow-up (e.g. fracture clinic appointment), participants in the intervention group presented again a trend for higher levels of satisfaction compared to the control group. One notable difference between groups was in the information provided to participants in both groups, with participants in the intervention group reporting higher levels of satisfaction with the information received. The superior satisfaction levels in participants randomised to the intervention group seemed to be associated with the downstream effect of immediate MRI, i.e. its ability to reach a quicker and definite diagnosis and, despite the increase in time in the ED, the sense of better understanding their clinical condition and its clinical management. Finally, the trial experience in patients randomised to the intervention group was associated with a statistically significant higher perception that the trial improved their clinical care.

Immobilisation time and time off work and informal care. It was hypothesised that the use of immediate MRI would lead to a lower number of participants being immobilised with plaster casts. This would then lead to a lower proportion of days off work or informal care. This hypothesis was not corroborated by the SMaRT trial as the intervention was associated with the detection of a higher proportion of clinically relevant injuries that required immobilisation with plaster cast. Nonetheless, time off work or informal care due to the suspected scaphoid fracture episode was not statistically different between groups. This seemed be due to the fact that participants with no fracture were immobilised with splint only (and not plaster cast) and still required time off work. Previous economic evidence, particularly economic modelling evidence, did not consider that symptomatic patients with no fractures could require time off work (Rua et al. 2017).

3.4.3 Sensitivity analyses

Sensitivity analyses. Deterministic sensitivity analyses around several parameters were considered. Regardless of the variations considered in all scenarios, the mean cost of the intervention group was always lower than the control group, i.e. the intervention led to cost savings although whether this cost-difference might be or not statistically significant varied. Moreover, in all 6-month cost-utility analyses, the intervention dominated the control group.

The first three sensitivity analyses considered changes to unit costs associated with the economic model. The increase in the unit cost of the immediate wrist MRI in the ED decreased, as expected, the cost savings associated with the intervention. In contrast, the increase in the unit cost of the fracture clinic appointments led to greater cost savings. This is explained by the fact that the intervention decreased the use of fracture clinic appointments, and thus, any increase in unit costs would only generate further cost savings. Variation to the assumed cost, the 'Did not attend' (DNA), was also subjected to sensitivity analysis. Any increase in the DNA cost led to an increase in the mean cost difference between the groups. The latter was due to the fact that the control group presented a higher number of missed appointments, particularly fracture clinic appointments. Given that a higher proportion of participants in the control group have formal follow-up at secondary care, it is more likely for participants in this group not to attend their appointments given that the wrist pain might subside between the day of the injury and the fracture clinic appointment (usually scheduled 7 to 10 days post-injury).

The analysis of existing reimbursement strategies was also considered given the potential impact on the financial sustainability of incorporating the intervention as part of routine care. This scenario led to a marginal increase in cost savings at both 3 and 6 months. However, the existing reimbursement strategy might lead to conflicting incentives as the use of immediate acute MRI is not currently incentivised given that ED episodes are reimbursed as bundle payments. Even though the use of immediate MRI is likely to increase the ED episode's tariff, this increase will not be enough for the provider to recoup the costs associated with the provision of MRI. Hence, in order to align the incentives of individual healthcare providers and the NHS, discussions with Clinical Commissioning Groups (CCGs) are required (see Chapter 7 for the detailed implementation plan).

Lastly, a broader perspective of analysis was also taken. From a societal perspective, healthcare costs that occurred in the private sector or non-healthcare costs, such as time off work costs, were included. Assuming a societal perspective, a higher mean cost difference between groups was estimated compared to the healthcare payer perspective (-£754 vs -£266). This finding was aligned with established economic literature. The systematic literature review conducted by the student found that, from a societal point of view, interventions that used advanced imaging were associated with higher cost savings (Rua et al. 2017). However, the cost difference between groups from a societal perspective in the SMaRT trial were not statistically significant (p=0.101). The latter seemed to be due to high attrition rates (54%) associated with self-reported resource use diaries, the instrument used to capture time off work. Hence, this constituted an important limitation to the trial's findings from a societal perspective.

3.4.4 Implementation plan: From research to clinical practice

Backed back by the SMaRT trial evidence, the next phase consisted of the implementation stage (Phase 4 in Figure 1). Three steps comprised the implementation plan to bridge the gap between research and real-world practice and to incorporate innovation into routine clinical care (Figure 44).

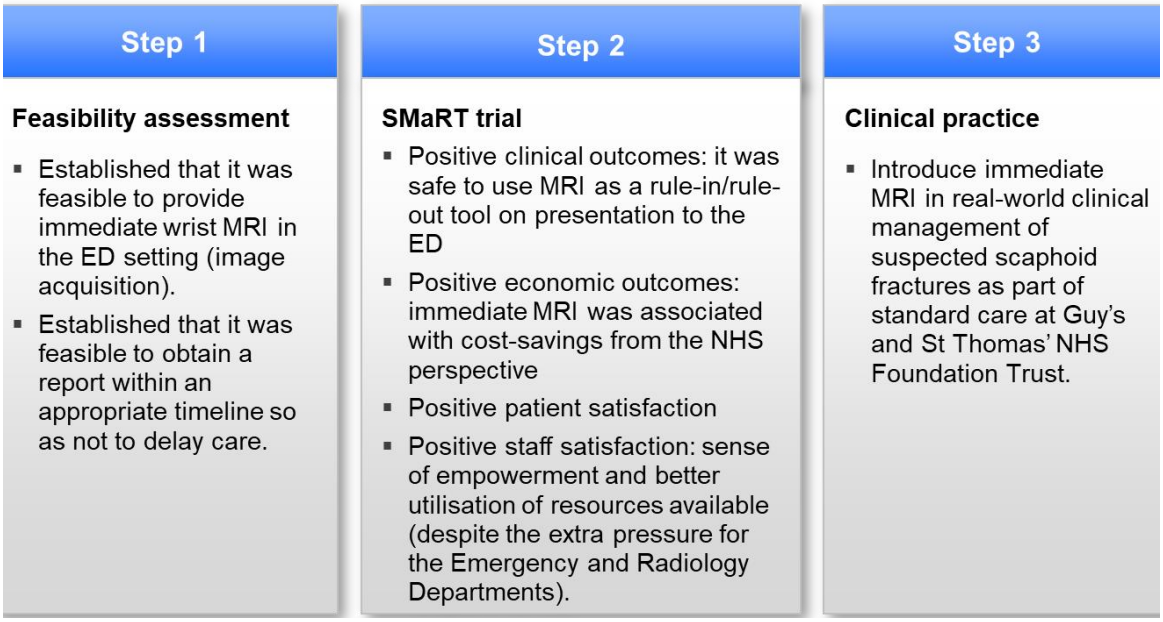


Figure 44. Three phases in the evolution of the project.

Step 1 comprised a feasibility assessment, of whether provision of immediate MRI in an acute setting was feasible. Out of 68 randomised to the MRI group, 63 (93%) received the intervention allocation. Five participants did not receive the intervention (i.e. MRI) due to: event of unforeseen claustrophobia (n=2), MRI not available within 1 hour (n=1), MRI not working (n=1) and presence of cochlear implant not mentioned during the initial screening process (n=1). Hence, despite the disruptive nature of using MRI in the ED, it was not feasible due to operational constraints only in 2.9% of cases (2 out of 68 participants).

Step 2 comprised the clinical and economic evaluation of the intervention. As summarised in this chapter, the intervention simultaneously generated cost savings and improved clinical outcomes and satisfaction levels. Hence, the first two phases established that the intervention was not only feasible but also associated with improved clinical and economic outcomes.

Step 3 emphasised incorporating the intervention as part of routine clinical practice at GSTT. Operational constraints associated with the provision of MRI for patients presenting to the ED at any time, as opposed to the MRI operating hours within the SMaRT trial eligibility criteria, were considered. Given the limited availability of MRI (not a 24/7 service), two options were contemplated, either: (a) keep the trial's inclusion criteria as the operating hours; or (b) roll-out the pathway for all patients. In the interest of equality and consistency across the patient pathway, it was decided that all patients presenting to the ED with a suspected scaphoid fracture should have immediate MRI. This posed some issues with the provision of MRI for patients presenting outside normal MRI working hours. In order to operationalise the clinical pathway, patients presenting to

the ED outside normal MRI working hours (e.g. 5 a.m.) were given the next available MRI slot (early morning slots) and discharged to the orthopaedic day care centre. If patients opted not to come back for the MRI scan, standard care with a subsequent fracture clinic follow-up was provided as per standard care prior to the introduction of immediate MRI (i.e. equivalent to the control group in the SMaRT trial). Figure 45 illustrates the clinical scenarios.

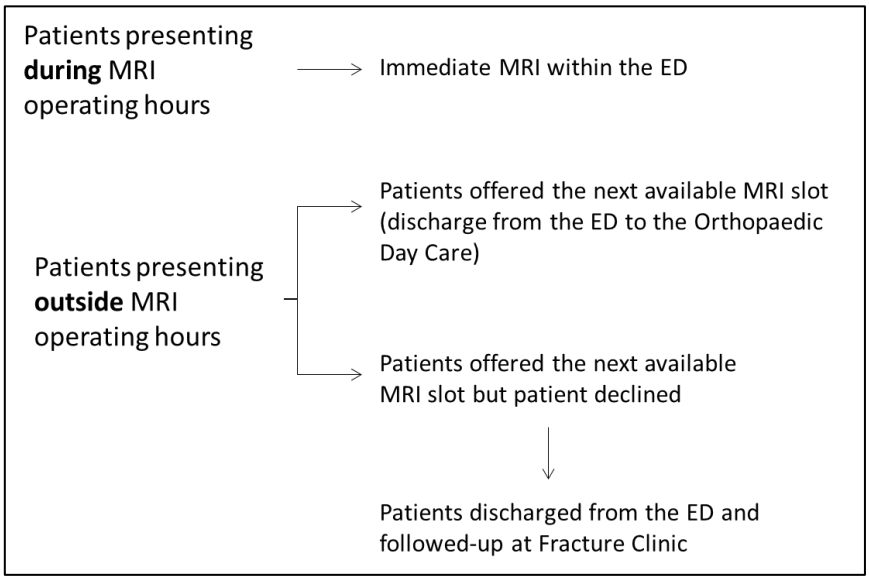


Figure 45. High-level description of the operational pathways considered in the provision of MRI based on the time patients presented to the ED.

Chapter 7 provides an in-depth discussion on the implementation plan to make immediate MRI part of routine care for all patients presenting at the ED with suspected scaphoid fracture.

3.4.5 Strengths, limitations and implications

Strengths:

The trial's pragmatic design aimed to assess how a clinical intervention works in a real-world clinical setting, particularly in the context of complex clinical settings (MacPherson 2004; S. Ramsey et al. 2005; Whicher et al. 2015). The ED is one such clinical setting. The strength of pragmatic clinical trials reside on their ability to impact on clinical practice. The positive result in the SMaRT trial provided evidence that the intervention was effective in real-world clinical practice. This contrasts with explanatory trials that are designed, not to inform decision makers, but to estimate a treatment's efficacy (MacPherson 2004; S. Ramsey et al. 2005; Whicher et al. 2015).

Several design features considered to minimise potential sources of biases and maximise were the trial results generalisability. First, the research question was clearly defined to assess the efficiency and effectiveness of immediate acute MRI in the management of suspected scaphoid fractures. Second, the inclusion criteria broadly reflected the heterogeneous population undergoing evaluation due to a suspected scaphoid fracture. Hence, participants recruited were likely to be representative of the wider population, contributing to the generalisability of the results. Third, the

comparator group was clearly defined as the standard care with the use of conventional radiograph only as part of the acute diagnostic pathway. This made the comparison between groups feasible as there was a clear difference between the control and the intervention groups. Fourth, the *a priori* sample size estimate was based on a clinical pilot and incorporated the heterogeneous mix of population and the probability of losses to follow-up (3 months for the primary outcome). Fifth, the randomised block design minimised biases, particularly allocation and selection biases. Sixth, all analyses were conducted on an intention-to-treat basis. Given that the intervention was designed as an add-on test (i.e. in addition to the control group), participants could not crossover between groups. This is commonly a potential disadvantage associated with other pragmatic trials (MacPherson 2004). Seventh, the selected outcomes were meaningful to everyday clinical practice and covered a wide range of dimensions, from efficiency to clinical. Eighth, the choice of final health outcomes (e.g. successful treatment of fractures, self-perceived health, costs) rather than intermediate outcomes (e.g. diagnostic accuracy) contributed to the decision around resource allocation (Drummond et al. 2004). Ninth, data-collection methods combined routinely collected data with research related data retrieved from an electronic case report form (eCRF). This approach contributed to a balance between data quality and interference with clinical practice (Meinecke et al. 2017). Tenth, the use of multiple clinical and financial databases, supplemented by primary care data provided by GPs, meant that there were no missing data for the primary outcome. This comprehensive approach guaranteed that any scaphoid-related NHS event was costed regardless of the healthcare provider or their location.

Limitations:

The first limitation derived from the SMaRT trial's single-centre design. To mitigate this limitation, the inclusion criteria reflected the heterogeneous population undergoing evaluation due to suspected scaphoid fractures across the UK.

The trial's exclusion criteria reflected the intervention's operational challenges. This led to intrinsic trial limitations. Most of the exclusion criteria were associated with the provision of MRI and subsequent reporting, as the trial only included participants that presented during MRI opening hours and the radiologist's services. This impacted the trial generalisability as findings from this trial might not be directly transferrable to different operating hours. In order to minimise this limitation and investigate the feasibility and transferability to other NHS-based hospitals, particularly outside normal working hours, the SMaRT trial led to another study aimed at evaluating the potential first-line use of radiographers to rule in or rule out suspected scaphoid fractures using MRI images on presentation to the ED.

Another important trial limitation was the lack of blinding. Given the clear differences between the two imaging modalities (i.e. conventional radiograph and MRI) it was not deemed possible to blind participants or staff to study allocation. However necessary, this constituted a trial limitation that could lead to conscious or unconscious bias from the participant and/or the routine care team staff. In order to mitigate this bias and given that had the potential to affect the primary outcome (e.g.

over or underutilisation of NHS resources), clinical pathways were disseminated and followed by clinicians.

Although the use of the EQ-5D-5L questionnaire, a non-disease specific instrument, is arguably not the best instrument to capture differences of effect between the groups given the injury's nature, we have opted to use it to follow the methodology suggested by the NICE.

Implications for Further Research:

The feasibility of providing an acute 24-hour service with MRI to diagnose suspected scaphoid fractures should be further evaluated in different clinical settings, from secondary to tertiary centres. Furthermore, particularly for healthcare providers that do not have MRI availability, the use of CT or dedicated extremity MRI scanners should be considered. Additionally, lessons learned from the design and implementation of the SMaRT trial are likely to provide a valuable insights to other researchers designing studies in an acute setting.

Implications for Policy and Clinical Practice:

The SMaRT trial was designed to provide evidence on which to base a decision. The results from the trial have shown that the immediate use of MRI in the management of scaphoid fractures was associated with a decrease in total costs and was cost-effective from a healthcare payer perspective. Furthermore, the intervention was linked with improved diagnostic accuracy, clinical outcomes and self-reported quality of life. Following on the evidence obtained in this trial, the use of immediate acute MRI in the ED is now used as an add-on test in the ED for the management of suspected scaphoid fractures with negative initial radiographs at GSTT.

3.5 Conclusion

The SMaRT trial evaluated the acute use of immediate MRI in the management of suspected scaphoid fractures. Given the intervention's innovative nature and the trial's pragmatic design, this trial provided empirical clinical and economic evidence on which to base future clinical practice.

The findings from the SMaRT trial showed that the use of immediate MRI led to both improved clinical and self-reported patient satisfaction outcomes whilst being cost-effective and reducing overall NHS costs associated with the holistic management of the suspected scaphoid fracture.

As a corollary, and despite its limitations, the SMaRT trial contributed to the field of knowledge by providing solid evidence on which to base UK and international clinical practice in favour of the use of immediate MRI in the management of patients with suspected scaphoid fractures. Since 28th October 2019, current standard care at GSTT has included the use of immediate MRI in the acute management of suspected scaphoid fractures with negative initial radiographs.

Chapter 4. Use of advanced imaging in the management of chronic headache

4.1 Introduction

4.1.1 Headache

Headache is the most common symptom reported in the community, affecting more than 90% of the population at some point in their lifetime, particularly women (Silberstein and Lipton 1993; Rasmussen et al. 1991). Most headaches are primary headache disorders, such as migraine, cluster or tension-type headaches, with secondary headaches, due to an underlying serious pathology (e.g. tumour, brain aneurysm), being far less common (Rasmussen et al. 1991). In fact, less than 0.1% of the lifetime prevalence of headache is associated with a life-threatening condition, which can include primary or secondary brain tumours (Symvoulakis et al. 2007, Department of Health 2012).

Headache is in the top ten international causes of disability (Stovner et al. 2007), with nearly half of sufferers reporting that it affects work, home or social activities (Boardman et al. 2003). A key study published in the Lancet estimated that tension-type headaches and migraines were, respectively, the second and third most common diseases in the world in both males and females, with migraine being responsible worldwide for about 3% of years lived with a disability (Vos et al. 2012; Leonardi and Raggi 2013).

4.1.2 The clinical challenge

Most people with headache self-manage, but over 4% of adults annually consult their General Practitioner (GP) complaining of headache (Latinovic, Gulliford, and Ridsdale 2006). GPs manage 97% of headache consulters, with 2% of these referred to neurologists and 1% to other specialists (Latinovic, Gulliford, and Ridsdale 2006).

The management of headaches depends on the quality of treatment provided in primary care, where this condition is predominantly managed (Ridsdale et al. 2008). Some patients with chronic headache tend to return frequently to primary care, concerned that the chronic headache might be a symptom of a serious underlying clinical condition, particularly a brain tumour. GPs acknowledge that they have made referrals for secondary care, both for a neurologist consultation or neuroimaging examination, in situations where they were unable to reassure patients (Morgan, Jenkins, and Ridsdale 2007). Morgan, Jenkins and Ridsdale (2007) and Ridsdale et al. (2007) found that referral for headache is often the outcome of patient pressure and anxiety interacting with GP characteristics, organisational factors and service availability rather than the disease severity itself. This contrasts with guidance from the National Institute for Health and Care Excellence (NICE), which does not recommend the use of neuroimaging for reassuring purposes (NICE, 2018). A US study estimated that a patient with new migraine headache or a flare-up chronic

headache had, respectively, a 39% (95% CI 24–54%) or 51% (95% CI 32–68%) probability of having neuroimaging routinely ordered even though guidelines specifically recommended against it (Callaghan et al. 2015).

From the neurologist's perspective, despite headache being mainly managed within primary care, was the most common cause for referral accounting for up to 22% of GP referrals to neurologists (Thomas et al. 2010). Similar numbers were found by Patterson and Esmonde (1993) and Ridsdale et al. (2011), who reported that between 20 to 30% of new referrals to neurologists were due to headache. Local estimates from King's College Hospital (KCH) and NHS Southwark suggested that headache accounts for 25% of local neurology appointments (Community Headache Service 2011). More recently, in 2018, evidence collected at GSTT showed that headache referrals increased over 10% annually between 2015 and 2018. The latter seemed to be due to growing patient demand for investigations and a lowering of GP referral thresholds.

In summary, despite the low level of referrals to secondary care (most patients are managed within primary care), the absolute number of headache episodes (due to its high prevalence) makes headache the most frequently listed reason for referral to the neurologist and thus eats up already severely constrained capacity. To inform future management of chronic headache, this study aimed to evaluate two existing pathways used in the management of patients with chronic headache: (i) referral to the neurology department; or (ii) direct access to neuroimaging, particularly MRI. These two pathways are illustrated in Figure 46 and described below.

4.1.3 Two pathways at Guy's and St Thomas' Hospital

Referral to a neurologist

In 2012/2013, it was estimated that GSTT received 672 referrals for headache, around 21% of the total number of neurology referrals (3,272). At KCH, over a period of 4 months in 2013, 178/357 (33%) of referrals were for headache (the equivalent of an annual estimate of 534 referrals). One third of referrals to neurology were from local GPs, one third from out of area, and the final third cross-referrals from other consultants, some of which were from the emergency department (Ridsdale et al., 2011).

In a study of reasons for attendance at GSTT's ED, the most common neurological reason was headache, accounting in 2006-7 for 1,565 ED visits per year (GSTT internal data). In 2018, the number of presentation to St Thomas' ED had increased over 50%, with an estimated number of over 2,500 ED visits with headache as the presenting complaint (GSTT internal information).

In summary, the management of headaches has led to a historical increase in the utilisation of secondary care resources, both acute and elective services.

Direct access to neuroimaging

In the absence of neurological signs, neuroimaging for chronic headache is not recommended by the US Headache Consortium Guidelines, the Scottish Intercollegiate Guidelines Network (Thomas

et al. 2010) or NICE (NICE 2018b). Nevertheless, brain scans – both CT and MRI – are commonly performed in patients with normal neurological examinations. As an example, a US study found that between 1995 and 2010, neuroimaging utilisation increased from 5.1% (95% CI, 2.7%-7.5%) to 14.7% (95% CI, 9.4%-20.0%) of all annual headache visits ($p<0.001$) (Callaghan et al. 2014).

The utilisation of advanced neuroimaging in the management of headache is due to its diagnostic accuracy, but moreover, the need to reassure patients that no serious underlying condition is causing the headache (Morgan, Jenkins, and Ridsdale, 2007). GPs described high levels of anxiety and patient pressure, predominantly caused by a fear of brain tumour as a cause for the headache referral (Morgan, Jenkins, and Ridsdale 2007; Elliot and Kernick 2011) although no statistical difference in neuroimaging findings between primary headache sufferers and healthy controls was found (0.58% vs 0.73%, $p>0.05$) (Wang et al. 2019).

In fact, GPs commonly acknowledge that they have made referrals to secondary care, both for a neurologist consultation or neuroimaging scan, in situations where they were unable to reassure patients (Morgan, Jenkins, and Ridsdale 2007). Scanning the patient's brain might provide reassurance for some patients but for others might not suffice. According to Thomas et al. (2010), the GP is in the best position to assess whether a negative or normal scan is likely to reassure the patient, decreasing the levels of anxiety and subsequent GP utilisation rates. Furthermore, an additional benefit from direct access to head MRI is to enable neurologists to better manage patients that end up being referred to them, allowing clinicians to concentrate on the patients' symptoms and headache management with the reassurance that no underlying structural pathology is present (Taylor et al. 2014).

An existing clinical pathway at GSTT has enabled primary care to directly access neuroimaging, particularly MRI (please see Figure 46). Overall head MRI utilisation for all types of referrals has increased over 120% in just four years, with 465 head MRI scans in 2015 compared to an estimated number of over 1,000 scans in 2019 (based on Jan-Sep 2019 departmental activity data). These data, however, included all scans in patients directly referred from primary care including, for example, patients with primary and secondary headaches. Despite the growth in referral rates, GPs pointed out that three aspects that could further increase referral rates: improved access to imaging; standardised reporting; and increased awareness about this pathway (Underwood, Kilner, and Ridsdale 2017).

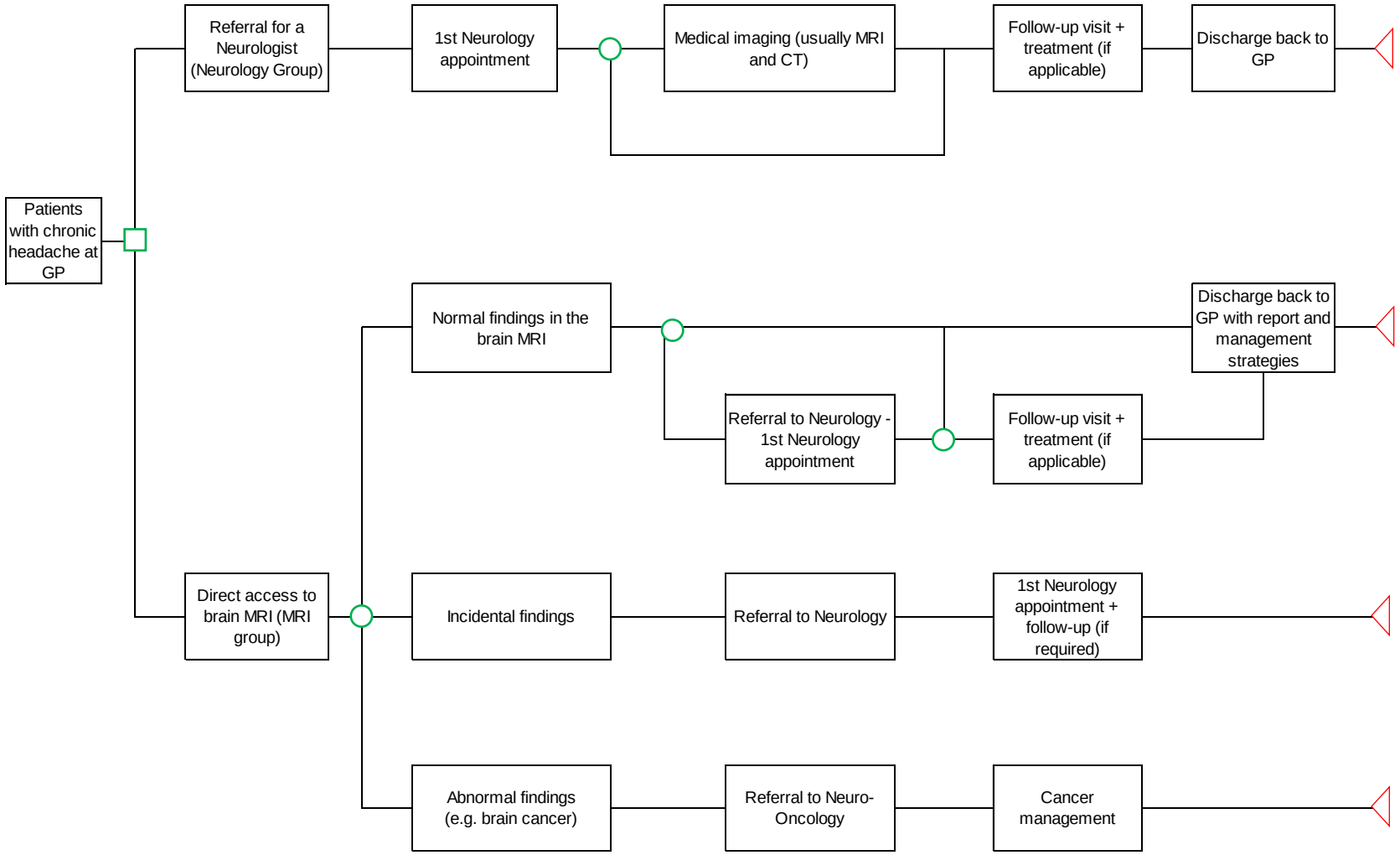


Figure 46. High-level illustration of two existing clinical pathways associated with the referral from GP due to chronic headache.

4.1.4 Economic evidence

A review of the literature assessed the existing economic evidence around the economic burden of headaches and the economic impact of using neuroimaging, particularly MRI, in the management of chronic headaches. Particular focus was given to evidence from studies conducted in the NHS.

Economic burden of headache

The Eurolight project (Linde et al. 2012) estimated the societal economic burden of headache in the European Union in 2008/09. Despite the benign nature of most headaches, the annual economic headache burden was estimated at 173€ billion, 111€ billion (64%) attributed to migraine, 37€ billion (21%) to medication-overuse, 21€ billion (12%) to tension-type headaches and the remaining 4€ billion (2%) to other types of headache (Linde et al. 2012). Although less prevalent, the economic burden of migraine was responsible for over half of the economic impact. This economic impact was divided in two components: (i) the actual direct cost associated with the consumption of healthcare resources; and (ii) the indirect costs due to absenteeism from work or lost productivity at work. Migraine or chronic migraine sufferers (>15 days/month) had more emergency department episodes, hospital visits and diagnostic tests than those with episodic migraine, leading to medical costs three times higher (Bloudek et al. 2012). Other studies, particularly in the US, although showing some variability, suggested that the vast majority of headache-related costs were indirect costs, particularly due to time off work (Hu et al. 1999; Hazard et al. 2009).

Two studies focused on the economic impact of headache in the UK: McCrone et al. (2011) and Osumili et al. (2018). McCrone et al. (2011), using 2003/04 data for patients presenting to primary care with headache, estimated a total cost of £956 million due to health service use and £4.8 billion including indirect costs, particularly lost productivity. Furthermore, this estimate was thought to represent an underestimate as many headache sufferers self-manage and do not consult their GP. In 2017, using 2012/14 data for patients presenting to headache specialists (either neurologists or GPs with special interest in headache), Osumili et al. (2018) estimated an overall £109 million and £850 million for health service use costs and total costs, respectively. In other words, McCrone et al. (2011) and Osumili et al. (2018) estimated that health service costs were only responsible for 20% and 13% of the total costs, respectively. Moreover, there were significant differences in terms of indirect costs considered in these two studies. Whilst McCrone et al. (2011) considered that most indirect costs derived from lost employment, Osumili et al. (2018) pointed out that informal care is responsible for 74% of total indirect costs with lost employment being responsible for only 13%. A potential explanation for this was that one third of participants were not employed due to other health issues and an additional 20% did not report any productivity loss (Osumili et al. 2018).

Impact of neuroimaging

This subsection chronologically summarises the evidence regarding direct access to imaging and the use of neuroimaging (either CT or MRI) as a direct alternative to management with a referral to a specialist, usually a neurologist.

A UK study at the Kent and Canterbury Hospital (data from 1994/95) evaluated direct access to MRI from primary care for different clinical reasons (Apthorp et al., 1998). A total of 170 MRI scans from 58 GPs were retrospectively evaluated. The authors found that where GPs had direct access to imaging, referrals to neurologists had been reduced, avoiding 41% of referrals to secondary care. In 24% of patients who were subsequently referred to secondary care, the specialty to which the patient was referred changed as a result of the MRI, thus improving referral appropriateness (Apthorp et al., 1998). Finally, 43% of GPs reported that direct access to MRI had changed their management patterns (Apthorp et al., 1998).

Howard et al. (2005) conducted a UK-based randomised controlled trial (participants recruited between 1999 and 2001, n=150), which showed that the use of neuroimaging in patients with chronic headaches had the potential to change patient management. The authors found that referral rates to neurology decreased from 23% in the control group compared to 1.3% when GPs had direct access to MRI. The cost implications of using direct imaging were estimated for two groups: Hospital Anxiety and Depression Scale (HADS) positive patients (n=66); and HADS negative patients (n=84). For patients with HADS positive findings, the authors found that the use of neuroimaging reduced overall 1-year costs (-£465, 95% CI: -£1028 to -£104) (Howard et al. 2005). The latter seemed to be due to participant's reassurance. However, the authors found that this reassurance effect seemed to be time limited, with participants being reassured at 3 months but not at 1 year post-recruitment.

Simpson et al. (2010) evaluated the direct use of CT from primary care for the management of chronic headaches in Greater Glasgow (data collected between 1999 and 2007, n=4,404). The authors aimed to evaluate the clinical and economic impact of direct access to neuroimaging. Out of 3,943 scans, 461 scans (10.5%) were reported as abnormal but in only 60 (1.4%) of patients were these findings potentially causative of the headache (i.e. non-incidental findings). Twenty-two brain tumours were identified in the entire cohort (0.5%). The authors estimated approximate cost savings of over £86,681 for the entire cohort (Simpson et al. 2010). It is relevant to highlight, however, that this estimate was not based on a rigorous resource use valuation methodology but rather a rough estimate based on GP's feedback. Furthermore, the authors described this as a cost-effectiveness analysis but only a cost analysis was reported.

Thomas et al. (2010) evaluated direct access to brain imaging for a subset of participants from the study by Simpson et al. (2010). This study was conducted in Tayside (Scotland) with an overall participation of 45% of the 309 local GPs. The utilisation of direct access to imaging, in this case CT, reduced referral rates in 86% of the cases during the follow-up period (average of 1.3 years per patient) (Thomas et al., 2010). The added number of CT scans performed through the open access service accounted for 4% of the annual CT scans performed across Tayside.

Kernick and Williams (2011) assessed the impact of providing GPs with direct access to neuroimaging for patients with headache. Kernick and Williams (2011) claimed that, although the yield for clinically significant findings in neuroimaging was below 1% (Tsushima and Endo 2005; Sempere et al. 2005), the value of any reassurance effect associated with neuroimaging remained

unknown (Howard et al. 2005; Kernick and Williams 2011). Furthermore, anxiety associated with incidental findings should not be disregarded (Kernick and Williams 2011; Taylor et al. 2014). In any case, the evidence suggested that neuroimaging affected the clinical pathway associated with the management of headache patients (Howard et al. 2005; Thomas et al. 2010; Simpson et al. 2010).

In addition, Kernick and Williams (2011) discussed the role of either CT or MRI in the management of patients with chronic headache. No decisive data were available to make a recommendation on the merits of CT or MRI, with the authors concluding that the utilisation of each modality was usually a function of its sensitivity, side effects, availability and costs (Kernick and Williams 2011). A more recent guideline, in 2019, from the British Society of Neuroradiologists, favoured the utilisation of MRI in a non-acute setting because of its greater accuracy to rule out secondary causes for headache (Good, 2019).

In summary, headache is responsible for a high economic burden across different healthcare systems and societies. The utilisation of advanced imaging is commonly used to reassure both clinicians and patients that no underlying health threatening condition is causing the headache. However, there is limited UK evidence around the economic impact of early advanced imaging in the management of chronic headache as well as the clinical management of incidental findings found in brain MRI or CT.

Can direct access to imaging ease bottlenecks?

The UK government is encouraging open access to scanning for different clinical conditions, particularly for suspected cancers (NICE 2016). The objective is to enable early diagnosis, improve accessibility to care and reduce pressure on specialist waiting lists and patient self-referral to the ED. This constitutes a major policy aim in a cash-strapped NHS.

The diagnosis and management of neurological diseases is particularly challenging in the UK compared to other western countries (Ridsdale et al. 2008). A major reason for this is the limited availability of neurologists, which is 5-10 times less than other comparable western countries (Ridsdale et al. 2008, Bateman 2011). If we consider that a 1% increase in the pattern of GP referrals to neurologists due to headache would almost double the demand for new neurology appointments at secondary care (Morgan, Jenkins, and Ridsdale 2007), the use of direct access to imaging holds great potential to release resources within the NHS. Some neurological conditions, such as epilepsy, are difficult for non-specialists to manage, and lack of primary to secondary care access may be a contributing factor in poor management and sudden death in epilepsy (Hanna et al. 2002). Freeing up much needed neurology capacity may improve patient management and reduce avoidable deaths.

Direct access to neuroimaging could be associated with a transfer of workload to radiology due to the increase in the number of referrals to MRI examinations. This is likely to be the case as not all patients directly referred to imaging from primary care would otherwise undergo an MRI examination as part of the management of their chronic headache. In addition, the workload

associated with the downstream management of incidental findings from neuroimaging remains unclear.

In summary, there is conflicting evidence surrounding the role of advanced imaging in the management of headaches, particularly with providing GPs with direct access to head MRI. Some authors (e.g. Wang et al. 2019) and regulatory agencies (e.g. NICE) do not recommend the utilisation of MRI for reassurance purposes in patients without neurological red flags. This decision is based on the low yield of relevant clinical findings in MRI among patients with headache. In contrast, other authors (e.g. Howard et al. 2005) considered that the role of advanced imaging in the management of headaches should be extended because, despite the low probability of finding relevant clinical findings, the reassurance of a normal head MRI to both referrers and patients improves clinical management and decreases the overall costs to the healthcare system. These benefits would justify providing GPs with direct access to head MRI.

4.1.5 Patient and NHS benefits and disbenefits

The underlying hypothesis is that the early use of a more expensive diagnostic tool (MRI) will reassure both patients and GPs that no serious underlying cause is present. Ultimately this might lead to:

- **An overall reduction of costs within:**
  - Primary care as patients with chronic headache will tend to: (i) attend less often to GP outpatient appointments; and (ii) reduce the use of medication.
  - Secondary care as patients directly referred to neuroimaging are more likely to have a lower number of secondary care contacts, particularly within the neurology department (reassurance effect).
- **An improvement of patient and GP satisfaction**:
  - From the patients' perspective - direct access to imaging is likely to: (i) improve access to care by decreasing overall waiting times; (ii) avoid unnecessary appointments and follow-up exams; and (iii) decrease overall pain and discomfort associated with the headache.
  - From the GP's perspective - direct access to imaging is likely to: (i) decrease the amount of time elapsed between referral and the initial secondary care appointment (compared to conventional referral to the neurology department); and (ii) improve the overall patient management strategy as GPs would have access to a standard MRI reports with guidance on how to proceed.

4.2 Methods

4.2.1 Aims, Objectives and Hypotheses

Aim of the study

The primary aim of the headache study was to evaluate whether direct GP access to MRI for patients with chronic headache was cost saving at 6 months compared to GP referral to the neurology department. The hypothesis was that direct access to MRI from primary care would reassure patients, decreasing the need for follow-up contacts and ultimately leading to lower total costs of care per patient associated with the management of headache compared to a referral to a neurology appointment. Hence, although MRI is an expensive imaging technology, a potential reduction of primary and secondary care utilisation - of GP and neurology appointments - might offset the added MRI cost. Furthermore, it is important to note that a proportion of patients would have had an MRI scan later during their pathway anyway, and so, for these patients, this would simply represent a shift in costs to an earlier time point within the clinical pathway rather than additional costs. This hypothesis was evaluated in multiple cost analyses.

Study objectives and hypotheses

One primary objective and five secondary objectives were considered.

Primary Objective: To estimate the 6-month costs associated with two existing clinical pathways in the management of chronic headaches: (i) referral from the GP to neurology services; or (ii) GP direct referral for a head MRI.

Null Hypothesis

There is no statistically significant difference between the 6-month cost per patient for patients with chronic headache referred from primary care to either: the neurology department; or directly to a head MRI.

Secondary Objectives:

Five secondary objectives were considered:

1. To estimate the 12-month costs following the initial episode at secondary care associated with direct access to MRI compared with referral to the neurology department.

Cost data up to 12 months of follow-up were used for the purposes of this secondary objective.

2. To perform cost-effectiveness analyses at 6 and 12-month following the initial episode at secondary care associated with direct access to MRI compared with referral to the neurology department.

Cost and utility data up to 12 months of follow-up were used for the purposes of this secondary objective. The cost-effectiveness analyses took into account both the costs and outcomes related to the intervention. Outcomes were expressed in QALYs derived from multiple EQ-5D-5L questionnaires.

3. To compare the levels of patient satisfaction associated with the two pathways: (i) referral from the GP to neurology services; or (ii) GP direct referral for a head MRI.

Patient satisfaction was evaluated using a non-standard questionnaire at 3 months following the initial episode at secondary care (either the initial neurology outpatient appointment or an MRI scan). The underlying hypothesis was that the use of MRI would improve the patient's overall satisfaction as it would decrease the overall number of contacts with secondary care and the time taken to obtain formal feedback/reassurance in each pathway.

4. To estimate the condition specific disability (using headache diaries and HIT-6 and MIDAS questionnaires and EQ-5D-5L questionnaire) associated with the two pathways: (i) referral from the GP to neurology services; or (ii) GP direct referral for a head MRI.

This objective evaluated the patient's self-reported quality of life and headache burden in both pathways using a generic questionnaire (EQ-5D-5L), two disease-specific questionnaires (HIT-6 and MIDAS) and headache diaries. The headache burden was based on the number of headache days and self-reported intensity. Data collected during the initial secondary care episode were used for baseline purposes.

5. To compare the time-off work due to chronic headache associated with the two pathways: (i) referral from the GP to neurology services; or (ii) GP direct referral for a head MRI.

The study assessed potential benefits from a broader societal perspective. Time off work due to chronic headache was recorded in headache diaries and questionnaires filled out quarterly.

4.2.2 Study design

The study was a single site, prospective, observational study. The study compared two existing clinical pathways in the management of patients with chronic headache with either GP referral to the neurology department or to direct access to head MRI. No changes to the existing clinical pathways were made as part of the study. The study was designed to inform future decisions around the management of patients with chronic headache.

Figure 47 presents the headache MRI study schema. The expected length of time for which each patient participated in the study was 12 months. Given the expected short-term impact of the intervention, it was considered that all relevant costs and outcomes were captured within the proposed 12-month follow-up. Data were collected quarterly (as highlighted in Figure 47 and Table 33).

Intervention groups

Participants were allocated to two groups: the neurology and the direct access to MRI groups. The allocation was decided *a priori*, by the referrer (in this case the GP) who decided which referral route would suit each participant. Subsequent care was consistent with standard care for each clinical pathway and is visually depicted in Figure 46.

Random allocation and blinding

No randomisation process took place.

The study was unblinded. Given the nature of the intervention and its impact on subsequent care, it was not deemed feasible to blind participants or research staff to the intervention. This lack of blinding may have led to potential conscious or unconscious performance bias, with differences in care received in both groups potentially affected by this feature. However necessary, this constituted a study limitation. The lack of blinding might also have led to attrition bias, i.e. different attrition rates in both groups. Different preventive steps were put in place to mitigate this risk, particularly: (a) ensuring good communication between participants and different members of the research team; (b) financial incentives for participants to comply with the study follow-up; (c) utilisation of databases that were not based on participants self-reporting data; and (d) the use of intention-to-treat analysis.

Follow-up Period

All participants were followed-up 12 months following study enrolment. Data were collected at baseline and follow-up data were collected quarterly, at 3, 6, 9 and 12 months post-recruitment.

4.2.3 Ethical Approval, Trial Registration and Funding

The Health Research Authority and Research Ethics Committee (West of Scotland – REC 4) approved the study research on 12th April 2016. The REC reference is 16/WS/0028 and the IRAS project ID is 163140.

The study started on 15th April 2016, with the first participant being recruited on 19th April 2016. The study was registered on clinicaltrials.gov (Clinical Trial Registration: NCT02753933) on 25th April 2016.

The headache study was fully funded by a grant secured from Guy's and St Thomas' Charity.

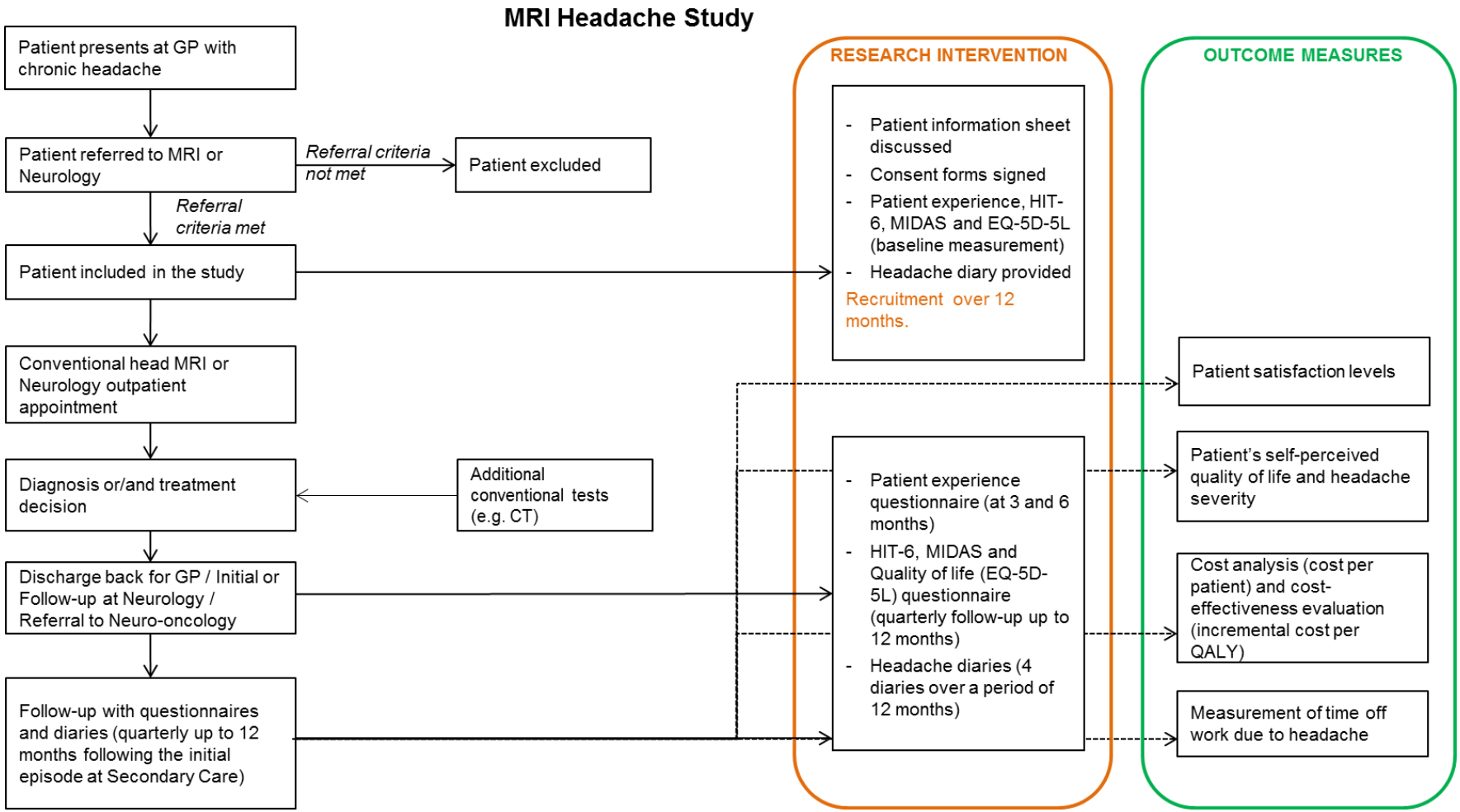


Figure 47. Structure of the headache study.

Table 33. Study flowchart for both study groups.

Activity	Timing of activity					
	Prior to registration for study	Immediately after registration into the study	After the initial episode at secondary care			
			3	6	9	12
Give patient information sheet, explain study and obtain signed informed consent	X					
Register patient demographics and clinical history		X				
Give patient registration pack (patient information sheet, copy of informed consent)		X				
Baseline questionnaires (HIT-6, MIDAS and EQ-5D-5L)		X				
Neurology appointment		X				
MRI scan		X				
Follow-up questionnaires (HIT-6, MIDAS and EQ-5D-5L)			X	X	X	X
Patient satisfaction questionnaire			X	X		
Headache diaries (4 diaries in total)		X	X	X	X	X

4.2.4 Selection, withdrawal of participants and sample size

Study Setting

Patients were recruited at GSTT, either at the neurology department or the radiology department.

Inclusion Criteria

Patients suitable for the study included every patient aged 16 years or over: (i) with chronic headache as a primary cause that lasted ≥ 15 days per month for more than 3 months; and (ii) referred from GP practices to GSTT, either directly referred to an MRI exam or a neurology appointment.

Exclusion Criteria

Patients were considered to be ineligible if at least one of the exclusion criteria was present:

- Children under the age of 16 years old;
- Patients with red flags as defined in the NICE guideline CG150 (NICE 2018b);
- Patients without chronic primary headache, i.e. a headache that did not persist for ≥ 15 days per month for more than 3 months;
- Patients with headache referred through the two week wait list;
- Patients who lacked capacity to give consent or participate in the study;
- Patient not fluent in English;
- Prisoners;
- Patients already taking part in a clinical trial of an investigational medicinal product.

Sample Size

The sample size estimate was calculated based on the primary endpoint, total 6-month healthcare costs. A total of 150 participants were needed in the neurology group and 99 participants in the MRI group in order to detect a cost difference of £300 assuming standard deviations of £750 and £500, respectively, with 85% power at the 5% two-sided significance level. The sample size was inflated by 20% to account for unknown cost distributions and attrition.

Losses to Follow-Up

If a patient moved from the GSTT catchment area, every effort was made to ensure the patient was followed up. If a patient was not contactable, individual patients' GPs were contacted to obtain information. If the participant requested to be withdrawn from the study, her/his data were excluded from the analysis. Given the mobility within the Greater London area and the relatively long follow-up (12 months), it was estimated that at least 10% of patients would be lost to follow-up.

End of Study

For regulatory purposes, the 'declaration of end of trial' form was submitted to ethical committees 12 months after recruitment of the final participant. The last participant (249th participant) was recruited on 22th February 2018.

4.2.5 Data Collection and outcomes at baseline and follow-up

Data were collected by a research team member at baseline and then quarterly up to 12 months following recruitment. Data collection was completed in February 2019.

Data at baseline were collected before the participant's initial appointment (either neurology appointment or MRI exam) at GSTT. Follow-up data were collected as per the participants' preference, either via phone, email or post. The participants' preference was established at baseline and recorded in the case report form.

Participant Demographics

A variety of information was captured at baseline, including:

1. Age;
2. Gender (male/female);
3. Ethnicity (e.g. White British, White Irish, White Other);
4. Employment status (e.g. full-time job, part-time job, wholly retired from work);
5. Qualification (e.g. high school, master's degree);
6. Presence of mental health condition (dichotomous variable);
7. Headache triggers (e.g. stress);
8. Chronic diseases (e.g. diabetes). This information was obtained from participants' self-reported data, data from primary care (classified under 'active problems') and secondary care clinic letters.
9. Quality of life questionnaires (headache specific: HIT-6 and MIDAS; non-headache specific: EQ-5D-5L).

At baseline, headache severity was estimated using headache diaries, HIT-6 and MIDAS questionnaires. In addition, data on primary care utilisation for the 12 months prior to the study recruitment were collected at baseline including how many times the patient used primary care resources (e.g. GP outpatient appointment) due to headache before referral to secondary care. Following referral from primary care, data on headache severity (using HIT-6, MIDAS and headache diaries) were also recorded on a quarterly basis post-recruitment.

Primary Outcome

The primary outcome was to estimate the 6-month costs associated with two existing clinical pathways in the management of chronic headaches: (i) referral from the GP to neurology services; or (ii) direct referral from the GP to a head MRI scan.

Perspective of Analysis

The study took a NHS perspective of analysis. Only costs of headache-related NHS events were considered. This approach was consistent with the methodology recommended by the NICE and other agencies (e.g. EUnetHTA) for the evaluation of interventions with potential impact on health outcomes (EUnetHTA 2015).

The estimate of the total costs from a NHS perspective was based on the multiplication of: (a) any headache-related healthcare events; by (b) the unit cost of such event.

Resource Use Measurement

Resource use data included contacts with NHS healthcare providers associated with the management of headache. The methodology used to measure resource use data was similar to the one described in the scaphoid chapter (see section 3.2).

Valuation of Unit Costs

For the purposes of the primary outcome, the valuation of unit costs was, whenever possible, based on NHS Reference Costs 2016-17 (NHS Improvement 2017). All secondary care contacts were costed using this strategy.

For primary care events, an average cost for appointment (e.g. GP face-to-face appointment, GP phone appointment) was derived from the Unit Costs of Health and Social Care 2016 and inflated to 2017 using the hospital & community health services (HCHS) index (Curtis and Burns 2017). The average GP face-to-face and phone appointments were estimated to be 9.22 and 23.4 minutes long, equivalent to, respectively, £36.50 and £118.10 (Curtis and Burns 2017). Nurse face-to-face and phone appointments were assumed to have the same time ratio as GP appointments, with an estimated cost of £19.50 and £8.00, respectively.

Medication costs were derived from the Prescription Cost Analysis database (NHS digital 2018) and estimated from clinical data, specifically secondary care clinic letters, information provided by primary care and participant self-reported data via the resource use diaries.

In the case of a 'Did Not Attend (DNA)' event, it was assumed that the NHS incurred a cost, estimated to be equivalent to 30% of the unit cost where the participant had attended. In any case, this assumption was subjected to deterministic sensitivity analyses to better understand the actual impact on the primary outcome.

Table 34 lists the valuation of all unit costs considered to estimate the primary outcome, including the reference and the rationale behind any assumption. Sensitivity analyses were performed and presented in subsection 4.2.6.

Table 34. Unit costs for primary and secondary care events considered in the headache study.

Category	Unit Type	Unit cost (£)	Reference
Primary care			
GP appointment (face-to-face)	Per appointment	£36.50	Unit Costs of Health and Social Care 2016 (Curtis and Burns 2017) and inflated to 2017 using the HCHS index.
GP home visit	Per appointment	£118.10	
GP phone appointment	Per appointment	£14.80	
Nurse appointment (face-to-face)	Per appointment	£19.50	
Nurse phone appointment	Per appointment	£8.00	
Secondary care			
Emergency department episode with headache as the presenting complaint	Per episode	£73 or £116	NHS Reference Costs 2017 (NHS Improvement 2017). Variable cost as was dependent on the type of episode, particularly the investigations performed.
Head MRI	Per scan	£146	NHS Reference Costs 2017 (NHS Improvement 2017).
Head CT	Per scan	£100	
Initial neurology appointment	Per appointment	£240	
Follow-up neurology appointment	Per appointment	£148	
Botox treatment	Per treatment	£238.80	Codes AB17Z: Nerve block or destruction of nerve, for pain management; Botox: Drugs – CCG only – no HRG code (NHS Improvement 2017).
Occipital nerve block	Per treatment	£635.90	
Headache-related inpatient episode	Per episode	£597.50	

Secondary Outcomes

This subsection summarises the methods used to collect data to estimate the five secondary outcomes.

1. *To estimate the 12-month costs following the initial episode at secondary care associated with direct access to MRI compared with referral to the neurology department.*

Cost analysis at 12 months followed the same principles and methods outlined for the primary outcome.

2. *To perform cost-effectiveness analyses at 6 and 12-month following the initial episode at secondary care associated with direct access to MRI compared with referral to the neurology department.*

The incremental analysis of effectiveness used QALYs as the measure of effect. QALYs are a generic measure of quality of life used to perform cost-utility analyses. A linear relationship was assumed between two time points (i.e. the QALYs between 0-3 months is the average between these two points). The cost-utility analysis, which is the preferred method of economic evaluation of NHS interventions, was performed according to NICE recommendations (NICE 2012a).

3. *To estimate and compare the levels of patient satisfaction associated with the two pathways.*

This secondary outcome was assessed via a questionnaire at 3 months post-recruitment. This questionnaire, based on a 5-point Likert scale, was used to assess any potential differences between the two groups. This questionnaire evaluated three dimensions: (i) referral process (time elapsed between referral from primary care to initial secondary care appointment); (ii) initial appointment (on the day satisfaction); and (iii) the overall experience three months after recruitment.

4. *To estimate the patient's self-perceived quality of life (using EQ-5D-5L questionnaire) and disease burden (using headache diaries and HIT-6 and MIDAS questionnaires) associated with both pathways.*

Quality of life was assessed using the EQ-5D-5L generic questionnaire. Similarly, the headache disability burden was assessed using two validated headache specific questionnaires, HIT-6 (Figure 48) and MIDAS (Figure 49). In both cases, these questionnaires were applied at baseline and quarterly up to 12 months post-recruitment. The HIT-6 questionnaire (Figure 48) measured the headache burden based on 6 questions, leading to a score range from 36 to 78. The higher the score, the higher the headache burden (Kosinski et al. 2003).

Figure 48. Illustration of the headache-specific HIT-6 questionnaire.

The MIDAS questionnaire (Figure 49) assessed the disability associated with the headache, ranging from 0 to 90. As with the HIT-6 questionnaire, the higher the score, the more severe the headache. Along with the headache diaries completed by participants, the MIDAS questionnaire collected the number of headache days per quarter (maximum of 90 days) and self-reported headache pain scores (ranging from 0, no pain at all, to 10, the worst pain ever) (Stewart et al. 2001).

Data retrieved from the MIDAS questionnaire were individually screened for errors by participants filling out the questionnaire. The total MIDAS score was determined by adding the scores of 5 questions. However, the sum of questions 1 and 2 and questions 3 and 4 could not exceed 90 days (mutually exclusive). Whenever needed the score from these questions was corrected by the

student in order not to exceed the 90 days ceiling. Similarly, whenever the number of headache days in the past 3 months exceeded 90 days, this was corrected to a maximum of 90 days.

The Migraine Disability Assessment Test

The **MIDAS** (Migraine Disability Assessment) questionnaire was put together to help you measure the impact your headaches have on your life. The information on this questionnaire is also helpful for your primary care provider to determine the level of pain and disability caused by your headaches and to find the best treatment for you.

INSTRUCTIONS

Please answer the following questions about ALL of the headaches you have had over the last 3 months. Select your answer in the box next to each question. Select zero if you did not have the activity in the last 3 months. Please take the completed form to your healthcare professional.

_____ 1. On how many days in the last 3 months did you miss work or school because of your headaches?

_____ 2. How many days in the last 3 months was your productivity at work or school reduced by half or more because of your headaches? (Do not include days you counted in question 1 where you missed work or school.)

_____ 3. On how many days in the last 3 months did you not do household work (such as housework, home repairs and maintenance, shopping, caring for children and relatives) because of your headaches?

_____ 4. How many days in the last 3 months was your productivity in household work reduced by half of more because of your headaches? (Do not include days you counted in question 3 where you did not do household work.)

_____ 5. On how many days in the last 3 months did you miss family, social or leisure activities because of your headaches?

_____ Total (Questions 1-5)

What your Physician will need to know about your headache:

_____ A. On how many days in the last 3 months did you have a headache? (If a headache lasted more than 1 day, count each day.)

_____ B. On a scale of 0 - 10, on average how painful were these headaches? (where 0=no pain at all, and 10= pain as bad as it can be.)

Scoring: After you have filled out this questionnaire, add the total number of days from questions 1-5 (ignore A and B).

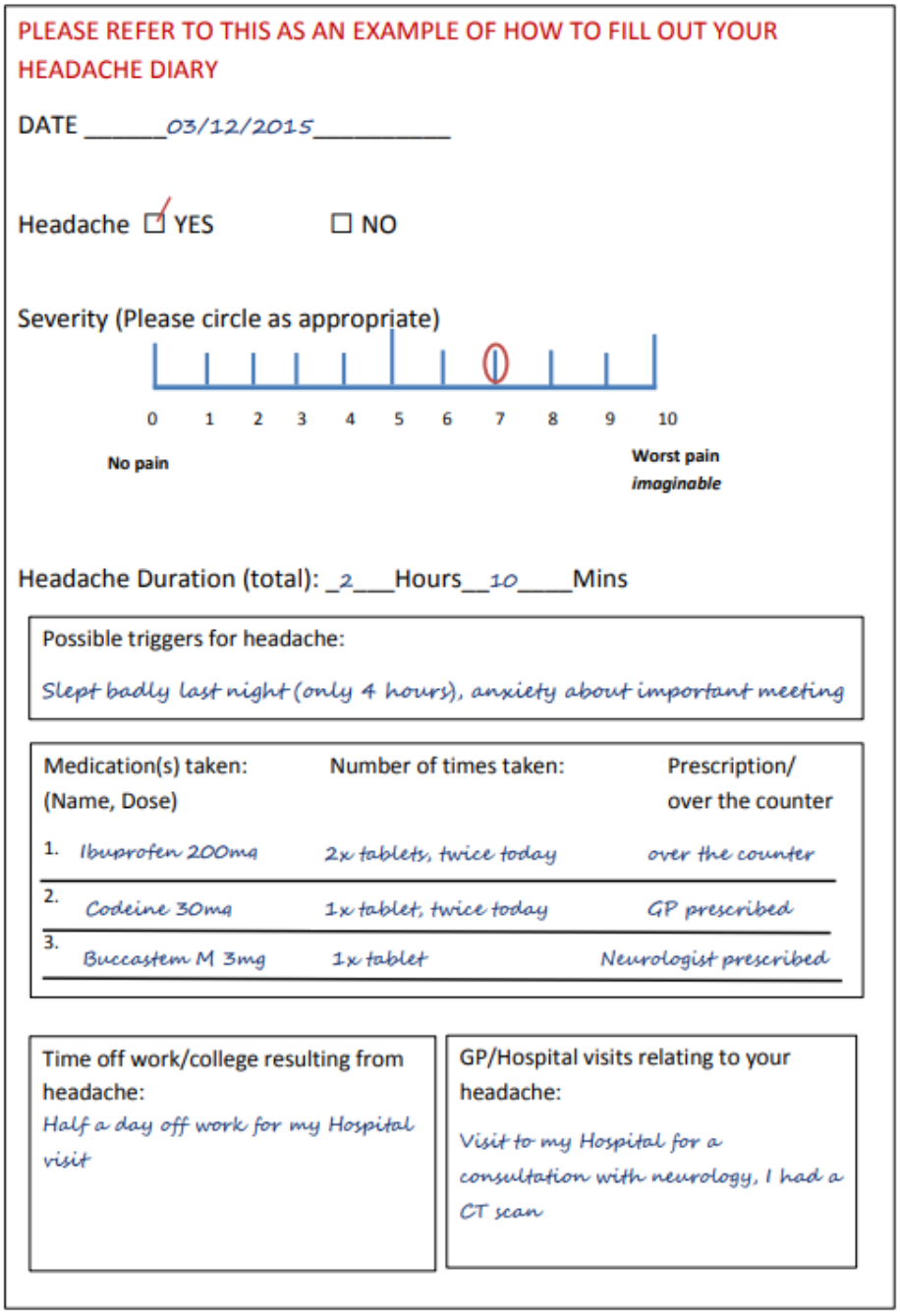
MIDAS Grade	Definition	MIDAS Score
I	Little or No Disability	0-5
II	Mild Disability	6-10
III	Moderate Disability	11-20
IV	Severe Disability	21+

Figure 49. Illustration of the headache-specific MIDAS questionnaire.

5. *To evaluate time-off work due to the chronic headache associated with the pathway with referral to the neurology department compared with direct access to MRI.*

Due to the anticipated attrition rate and potential for missing data, time off work due to chronic headache was collected using two sources of information: (i) participant diaries; and (ii) MIDAS questionnaire.

Participant diaries were either paper-based (Figure 50) or using a mobile app (Figure 51) as per participant's preference established during recruitment. This study involved Patient and Public Involvement (PPI) during the design phase, where prior to the start of the research, a group of patients suffering from chronic headache were consulted to co-produce the study mobile app. In these diaries, participants were asked to periodically record several variables, specifically: presence of headache (dichotomous Yes/No variable); headache duration; headache triggers; medication taken due to headache; time off work/college; and any healthcare visit due to headache. The study diary was based on existing headache diaries used as part of routine clinical practice.



PLEASE REFER TO THIS AS AN EXAMPLE OF HOW TO FILL OUT YOUR HEADACHE DIARY

DATE _____03/12/2015_____

Headache ☑ YES ☐ NO

Severity (Please circle as appropriate)

0 1 2 3 4 5 6 7 8 9 10

No pain — Worst pain imaginable

Headache Duration (total): _2___Hours__10____Mins

Possible triggers for headache:

Slept badly last night (only 4 hours), anxiety about important meeting

Medication(s) taken: (Name, Dose)	Number of times taken:	Prescription/ over the counter
1. Ibuprofen 200mg	2x tablets, twice today	over the counter
2. Codeine 30mg	1x tablet, twice today	GP prescribed
3. Buccastem M 3mg	1x tablet	Neurologist prescribed

Time off work/college resulting from headache:

Half a day off work for my Hospital visit

GP/Hospital visits relating to your headache:

Visit to my Hospital for a consultation with neurology, I had a CT scan

Figure 50. Print screen of the paper-based diary used in the headache study.

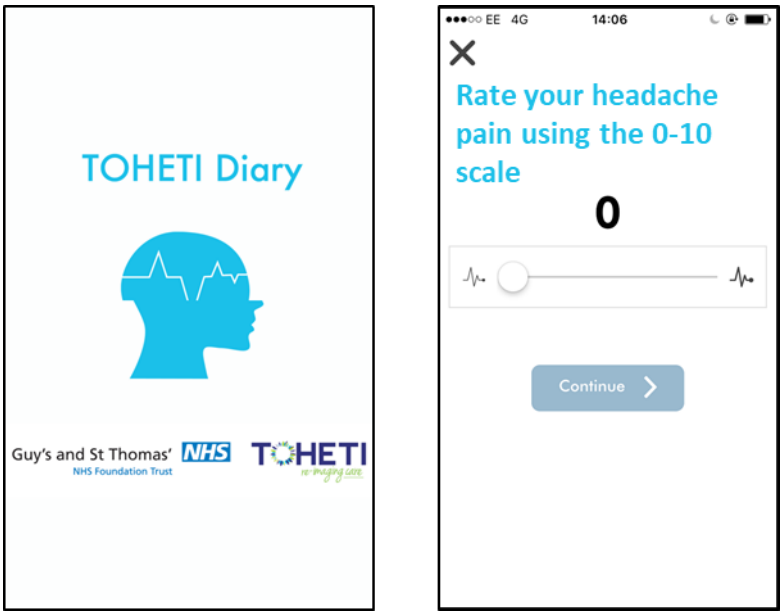


Figure 51. Print screens of the mobile diary app used in the headache study.

Two questions from the MIDAS questionnaire assessed the number of work days missed in the last 3 months and the number of days in the last 3 months where productivity was reduced by more than 50% (see Figure 49). Whenever there were discrepancies between the two documents, data captured via the MIDAS questionnaire were preferred because they were more likely to reflect the period of time considered. No information about participants' earnings was collected and hence the economic cost of lost productivity was estimated by multiplying the number of lost work days with the average national wage rate (Office for National Statistics 2019b).

Data on time off work were used to evaluate potential differences between groups and ultimately estimate potential implications from a societal perspective (as a secondary analysis).

4.2.6 Statistical Analyses

Analysis Population

All analyses were based on the principle of intention-to-treat, i.e. all participants recruited were analysed as per the study group allocation, regardless of whether they actually received the intended treatment, any protocol deviations or potential losses to follow-up (Peacock, Kerry, and Balise 2017).

Data Cleaning and Data Validation

All baseline and follow-up data cleaning was performed prior to any data analysis.

Baseline data were captured via a paper-based case report form during recruitment and then added to a web-based form. During this process, the PhD student screened the data for inconsistencies. For any potential data errors in the original hand-written data packs, participants or members of the research team were asked for clarification (e.g. date of birth and age do not match) and amendments were made to the original data set (data editing).

The NHS resource use measurement constituting the primary outcome was obtained by merging six different sources of information, grouped in two areas: medical records; and data self-reported by participants. With regards to medical records, multiple primary and secondary care databases were used, particularly: (1) Electronic Patient Record (EPR) and (2) Patient Information Management System (PIMS), used to map the elective pathway associated with the management of chronic headaches; (3) Symphony, an Emergency Department software used to map the acute component of the pathway (e.g. attendances to ED with headache as presenting complaint; (4) Computerised Radiology Information System (CRIS), used to capture any data related to diagnostic scans (e.g. CT, MRI); and (5) primary care data, allowing the mapping of any healthcare event that happened outside the GSTT remit (e.g. GP appointments, attendances to ED in other hospitals). In addition, data self-reported by participants supplemented this information, providing utilisation data, particularly any activity that happened within primary care or at any other secondary care provider other than GSTT or even in the private sector. This comprehensive data collection methods (already detailed in subsection 3.2 and Figure 33) considered the validation of data using multiple datasets.

Missing Data

Participants were not excluded from the analysis due to missing data, particularly data related to the primary outcome. Only data from participants that formally withdrew informed consent excluded from the analyses. A very high degree of costing data completeness was expected given the comprehensive data collection methodology.

Baseline comparability of groups

Continuous data were summarised by: frequency, mean, standard deviation, minimum, first and third quartile, median and maximum. Tabulations of frequencies for categorical data were presented, as well as the percentage (%) relative to number of non-missing values within the respective intervention group, unless otherwise specified.

Given the non-randomised study design, significance testing was performed on the baseline variables between intervention groups. Chi-square tests were used to assess categorical variables. Quantitative variables were tested for normality using the Shapiro-Wilk test and, depending on this result, independent t-test or Mann-Whitney U test analyses were performed. The Levene's test was used to assess the homogeneity of variance [consistent with Peacock, Kerry, and Balise (2017)]. A p-value of $p<0.05$ was considered as statistically significant.

Cost Analyses and Economic Evaluation

Primary Objective:

To estimate the 6-month costs associated with two existing clinical pathways in the management of chronic headaches: (i) referral from the GP to the neurology services; or (ii) direct referral from the GP to the MRI services.

Total costs were calculated based on the multiplication of any headache-related healthcare events by the unit cost of each event. Resource use data included contacts with any NHS healthcare

provider associated with the management of chronic headache. These included, among others, visits to GPs or headache clinical nurse specialist, inpatient care, neurologist or other headache-related outpatient visits (e.g. psychiatry), physiotherapist, visits to the emergency department, advanced imaging such as CT and MRI.

First, given the skewness associated with cost distributions (Peacock, Kerry, and Balise 2017), all cost differences between groups were assessed using GLM with an identity-link and gamma distribution. An identity link function instead of a log link was used in order to avoid potential biases (Polgreen and Brooks 2012; Barber and Thompson 2000; Peacock, Kerry, and Balise 2017). In the identity function covariates provide an additive effect on the mean (Jones 2011), allowing a direct interpretation of coefficients due to the lack of data transformation (as opposed to the log link function). A gamma distribution assumes a variance proportional to the square of the mean (Jones 2011). Continuous data such as healthcare costs, with skewed distribution and variation that increases with the mean, can be modelled with a gamma distribution (Peacock, Kerry, and Balise 2017). An unadjusted GLM cost analysis with the study group (neurology group vs MRI group) as only covariate was performed as the first step.

Second, given the study's observational design, the two groups being compared can be different due to the lack of randomisation (Manca and Austin 2008; Moran et al. 2007; Jones 2011; Cepeda 2003). This might lead to the presence of confounders with potential impact on the outcome being assessed. Potential differences in baseline characteristics that might confound the primary outcome were assessed. Such baseline characteristics included: (i) demographic (age, gender, qualification); (ii) clinical baseline characteristics [number of headache days, severity of headache pain, number of chronic conditions, self-perceived quality of life (EQ-5D-5L), anxiety and headache-specific questionnaire scores (HIT-6, MIDAS)]; and (iii) resource use utilisation prior to recruitment (number of GP appointments in the last 12 months prior to referral). All these baseline variables were assessed for statistical differences at baseline using a conservative threshold ($p<0.10$) and, if statistically different, were included as covariates in a second GLM (adjusted GLM). For all GLM analyses, group difference estimates and associated confidence intervals were reported, together with p-values.

An alternative statistical method, bootstrap models were considered for the analysis and compared against GLM (Gray et al. 2011). Bootstrap is a robust approach for the analysis of skewed cost data and can be recommended as either a secondary analysis to check the use of standard parametric methods or used as the primary statistical analysis (Barber and Thompson 2000).

All analyses were performed using the software Stata 15.0.

<u>Secondary Objectives:</u>

1. *To estimate the 12-month costs following the initial episode at secondary care associated with direct access to MRI compared with referral to the neurology department.*

As with the primary outcome, both GLM and bootstrap methods were used for the 12-month cost analysis.

2. *To perform cost-effectiveness analyses at 6 and 12-month following the initial episode at secondary care associated with direct access to MRI compared with referral to the neurology department.*

Cost-effectiveness analyses at 6 and 12-month were performed using bootstrap models with QALYs as the measure of effect (Briggs, Claxton, and Sculpher 2011). QALYs were calculated from utility scores at 6 and 12 months derived from the use of EQ-5D-5L questionnaire (Devlin et al. 2018) at five points in time: baseline (month 0), 3, 6, 9 and 12-month using area under the curve methods assuming linear movement between adjacent time points (Drummond et al. 2004). If utility data were missing, multiple imputation methods were used to assess the assumption that the data were missing at random. Missing data were imputed using 'multiple imputation using chained equations' (MICE), with the number of multiply imputed data-sets to be equal to the fraction of incomplete service-use information (White, Royston, and Wood 2011). One thousand bootstrapped replicates of differences in costs and outcomes were presented on cost-effectiveness planes. Cost-effectiveness acceptability curves showed the probability of the direct access to head MRI group being cost effective compared to the neurology group at varying thresholds of willingness to pay.

3. *To estimate and compare the levels of patient satisfaction associated with the two pathways.*

Satisfaction data were evaluated using Chi-square tests. A p-value of $p<0.05$ was considered as statistically significant.

4. *To estimate the patient's self-perceived quality of life (using EQ-5D-5L questionnaire) and disease burden (using headache diaries and HIT-6 and MIDAS questionnaires) associated with both pathways.*

Differences in the mean score for the three questionnaires (EQ-5D-5L, HIT-6 and MIDAS) were assessed for normality using the Shapiro-Wilk test. Depending on the findings, either independent t-test or Mann-Whitney U test were used to assess group differences for these variables.

5. *To evaluate time-off work due to the chronic headache associated with the pathway with referral to the neurology department compared with direct access to MRI.*

Differences in the mean pain score and time off-work were assessed for normality using the Shapiro-Wilk test. Depending on the findings, either independent t-test or Mann-Whitney U test were used to assess group differences for these variables.

Losses to Follow-Up

Potential statistically significant differences in losses to follow-up were compared between the two groups. The primary outcome estimate was adjusted using a GLM assuming a conservative threshold of significance for follow-up differences ($p<0.10$) between the two study groups (second adjusted GLM).

4.3 Results

4.3.1 Data Validation and Completeness

Only participants that withdrew the informed consent were considered lost to follow-up and hence not included in the data analysis.

Baseline data were complete except for some MIDAS questionnaires (missing data in four and five participants in the neurology and MRI group, respectively).

With regards to the primary outcome, data from primary and secondary care databases were, respectively, 97% complete (n=124 for the neurology group, n=92 for the MRI group) and 100% complete (n=128 for neurology group, n=95 for the MRI group). In the absence of data from both primary care databases and self-reported data, any resource use outside GSTT was missing. Missing values were imputed using the mean values from the respective group.

With regards to secondary outcomes, data from the EQ-5D-5L questionnaire at baseline and 6 months post-recruitment were, respectively, 99.6% complete (n=127 for the neurology group, n=95 for the MRI group) and 36% complete (n=55 for the neurology group, n=26 for the MRI group). Data from the HIT-6 questionnaire at baseline and 6 months post-recruitment were, respectively, 99% complete (n=128 for the neurology group, n=92 for the MRI group) and 21% complete (n=35 for the neurology group, n=12 for the MRI group). Data from the MIDAS questionnaire at baseline and 6 months post-recruitment were, respectively, 96% complete (n=124 for the neurology group, n=90 for the MRI group) and 27% complete (n=46 for the neurology group, n=15 for the MRI group).

4.3.2 Participant Flow

Participant flow associated with the headache study is illustrated in the diagram below (Figure 52). A total of 249 participants were recruited, 150 in the neurology group and 99 in the MRI group. A hundred per cent of participants recruited received the respective allocation treatment.

With regards to follow-up, 15% (n=22) and 4.0% (n=4) participants withdrew the informed consent in the neurology and MRI group, respectively, and hence were considered lost to follow-up. No other participants were considered lost to follow-up.

Participants who did not withdraw their informed consent were included in the analysis, equivalent to 128 participants (85% of the original sample size) and 95 (96%) participants in the neurology and MRI group, respectively. All analyses were based on an intention to treat principles.

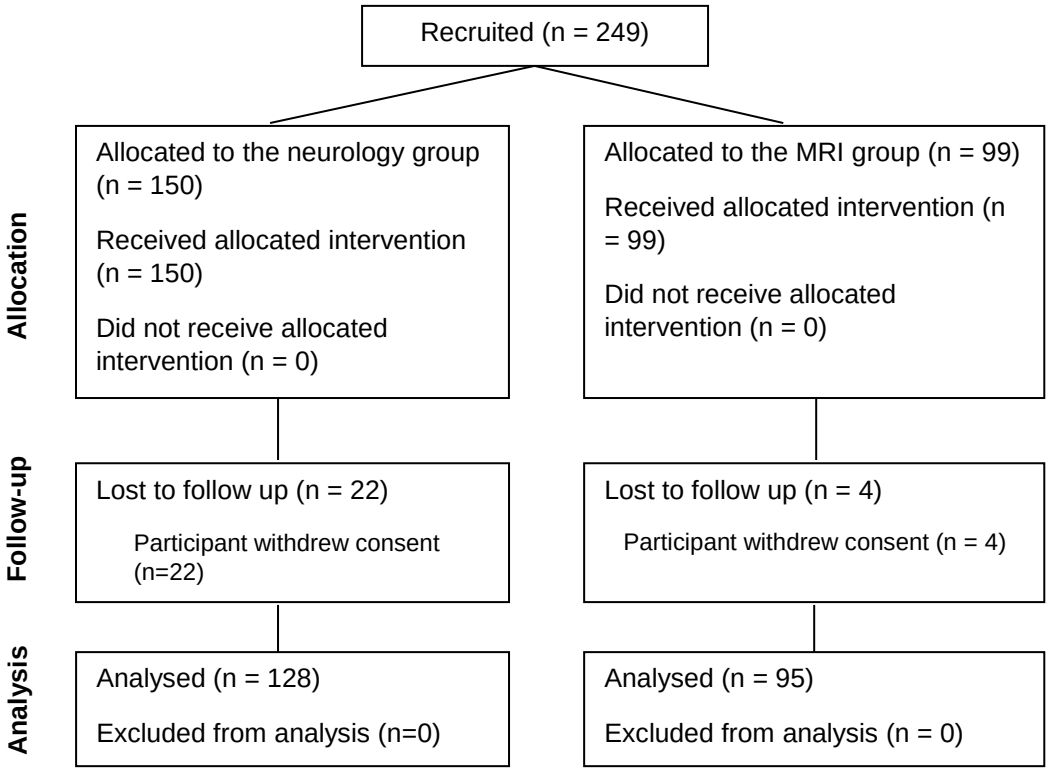


Figure 52. Participant flow chart for the headache study.

4.3.3 Participant Characteristics – Baseline

Table 35 and Table 37 (for categorical variables) and Table 36 (for continuous variables) report the baseline sociodemographic and outcome variables organised per study group. All participants who did not withdraw consent were included in the baseline analysis (n=223), distributed in the neurology group (n=128) and the MRI group (n=95). As documented in the *a priori* Statistical Analysis Plan (SAP), significance testing between intervention groups was performed given the non-randomised study design.

Categorical data

Table 35 describes the participants' characteristics (gender, ethnicity, employments status, qualification and mental health condition) organised per study group. All statistical analyses were based on Chi-square tests.

The large majority of participants were female, both in the neurology group (81%) and the MRI group (68%). This difference was statistically significant between groups ($\chi2(1)=4.26$, $p=0.039$).

The largest ethnic group was White British in both groups (42% and 39% in the neurology and the MRI group, respectively), followed by Black or Black British African (21% and 14% % in the neurology and the MRI group, respectively) and any other white background (8.6% and 22% in the neurology and the MRI group, respectively). The difference between groups approached statistical significance ($\chi2(13)=20.69$, $p=0.079$).

The majority of participants reported high school as the highest education level (45% and 35% in the neurology and MRI group), followed by Bachelor's degree (24% and 31% in the neurology and MRI group). The difference between groups was not statistically significant ($\chi2(6)=4.84$, $p=0.565$).

With regards to the employment status, most participants were employed full-time (39 and 38% in the neurology and MRI group), with 6.3% of participants in both groups being permanently sick or disabled. The difference between groups was not statistically significant ($\chi2(9)=4.95$, $p=0.839$).

In terms of psychiatric comorbidities, 17% of participants in the neurology and 12% in MRI group reported at least one mental health condition. The difference between groups was not statistically significant ($\chi2(1)=1.36$, $p=0.243$).

Continuous data

All quantitative variables were assessed for normality using the Shapiro-Wilk test, based on a significance value of $p<0.001$. As the assumption of normality was not met, the Mann-Whitney U test was used.

The mean age (SD) of participants was 38.4 (14.1) and 40.0 (14.6) years old in the neurology and the MRI group, respectively (see Table 36 for further detail). The difference between groups was not statistically significant ($p=0.514$).

The number of active health problems at baseline was estimated. The mean (SD) number of active health problems was 2.0 (1.5) and 1.8 (1.4) in the neurology and MRI group, respectively. The difference between groups was not statistically significant ($p=0.277$).

The mean (SD) number of headache triggers was estimated at 2.1 (1.8) and 1.8 (1.4) in the neurology and MRI group, respectively. The difference between groups was not statistically significant ($p=0.378$).

The mean number of days (SD) from referral to first appointment, either a neurology appointment or head MRI scan, was measured as a proxy of accessibility to care. The neurology and MRI group represented 110 (35) and 39 days (17), respectively, which was statistically significant ($p<0.001$). It was assumed that the MRI intervention would only be delivered once the actionable information, i.e. the MRI report, was available to the referrer (GP). The mean (SD) time elapsed from referral to MRI results being available to GP at 70 days (35). Deducting this with the time it took for patients to have the first appointment in the neurology group, the difference between groups in waiting times remained statistically significant ($p<0.001$).

Health-related quality of life (HRQL) at baseline was assessed using three questionnaires, one generic (EQ-5D-5L) and two headache-specific (HIT-6 and MIDAS). Again, it is relevant to highlight that all questionnaires were administered before the neurology appointment or the MRI scan. The mean utility (SD) at baseline (generated from the EQ-5D-5L questionnaire) was 0.809 (0.182) and 0.830 (0.195) for the neurology and MRI groups, respectively. This trend was found to approach statistical significance ($p=0.097$). The mean (SD) self-reported quality of life score, based on a 0-100 visual analogue scale, was estimated at 64.0 (18.8) and 70.8 (20.2) for the neurology and MRI groups, respectively. This difference was found to be statistically significant ($p=0.005$).

Table 35. Categorical sociodemographic and baseline characteristics of participants.

		Neurology		MRI		p-value
		N	%	N	%	
Gender	Male	25	20%	30	32%	p=0.039
	Female	103	81%	65	68%	
Ethnicity	White British	54	42%	37	39%	p=0.079
	White Irish	3	2.3%	2	2.1%	
	Any other White background	11	8.6%	21	22%	
	Mixed White and Black Caribbean	3	2.3%	4	4.2%	
	Mixed White and Asian	0	0.0%	2	2.1%	
	Any other Mixed background	2	1.6%	2	2.1%	
	Asian or Asian British Indian	3	2.3%	1	1.1%	
	Asian or Asian British Bangladeshi	1	0.8%	2	2.1%	
	Asian or Asian British Pakistani	3	2.3%	0	0.0%	
	Any other Asian background	6	4.7%	3	3.2%	
	Black or Black British Caribbean	10	7.8%	4	4.2%	
	Black or Black British African	27	21%	13	14%	
	Any other Black background	3	2.3%	0	0.0%	
	Any other ethnic background	2	1.6%	4	4.2%	
Qualification	Advanced Graduate Work or PhD	3	2.3%	4	4.2%	p=0.565
	Master's Degree	15	12%	12	13%	
	Bachelor's Degree	31	24%	29	31%	
	High School	57	45%	33	35%	
	Did not finish High School	8	6.3%	10	11%	
	Prefer not to answer	14	11%	7	7.4%	
Employment	Employee in full time job (30 hours or more a week)	50	39%	36	38%	p=0.839
	Employee in part-time job (under 30 hours a week)	18	14%	18	19%	
	Self-employed, full or part time	12	9.4%	8	8.4%	
	Full-time education at school, college or university	16	13%	9	9.5%	
	Doing something else	2	1.6%	2	2.1%	
	Permanently sick/ disabled	8	6.3%	6	6.3%	
	Looking after the home	4	3.1%	6	6.3%	
	Unemployed and available for work	12	9.4%	6	6.3%	
	Wholly retired from work	4	3.1%	4	4.2%	
	Prefer not to answer	2	1.6%	0	0.0%	
Mental Health condition	Yes	22	17%	11	12%	p=0.243
	No	106	83%	84	88%	

Table 36. Continuous sociodemographic and baseline characteristics of participants.

Variable		N	Mean	Standard Deviation	Minimum	Percentile 25	Median	Percentile 75	Maximum	p-value
Age	Neurology	128	38.4	14.1	16.0	25.0	37.0	51.0	79.0	p=0.514
	MRI	95	40.0	14.6	16.0	29.0	38.0	49.0	87.0	
Active Health Problems	Neurology	128	2.0	1.5	0	1.0	2.0	3.0	6.0	p=0.277
	MRI	95	1.8	1.4	0	1.0	2.0	3.0	6.0	
Number of headache triggers	Neurology	128	2.1	1.8	0	1.0	2.0	3.0	7.0	p=0.378
	MRI	95	1.8	1.4	0	1.0	2.0	3.0	8.0	
Days Referral to 1st appointment	Neurology	128	110	35	4	97	112	132	227	p<0.001
	MRI	95	39	17	2	36	41	46	90	
MRI: referral to results	MRI	94	70	35	7	46	76	94	206	p<0.001
EQ-5D-5L utility baseline	Neurology	128	0.809	0.182	-0.185	0.747	0.851	0.942	1.000	p=0.097
	MRI	95	0.830	0.195	-0.170	0.777	0.881	0.942	1.000	
EQ-5D-5L visual analogue scale (VAS) baseline	Neurology	128	64.0	18.8	10.0	50.0	70.0	80.0	100.0	p=0.005
	MRI	95	70.8	20.2	10.0	55.0	70.0	90.0	100.0	
HIT-6 baseline	Neurology	128	65.0	5.3	44.0	63.0	66.0	68.0	78.0	p=0.006
	MRI	95	62.6	7.3	36.0	59.0	63.0	68.0	78.0	
MIDAS baseline score	Neurology	128	57.8	54.0	0	17.0	42.0	79.0	215.0	p=0.075
	MRI	95	44.8	44.9	0	13.0	29.0	68.0	240.0	
MIDAS headache days	Neurology	128	51.6	31.5	.0	24.0	48.0	90.0	90.0	p=0.038
	MRI	95	42.8	30.7	1.0	15.0	36.0	70.0	90.0	
MIDAS headache pain score	Neurology	128	6.9	1.8	0	6.0	7.0	8.0	10.0	p=0.778
	MRI	95	6.9	1.9	2.0	6.0	7.0	8.0	10.0	

The mean (SD) HIT-6 score at baseline was estimated at 65.0 (5.3) and 62.6 (7.3) for the neurology and MRI group, respectively (Table 36). This difference was statistically significant (p=0.006). The mean (SD) MIDAS score at baseline was estimated at 57.8 (54.0) and 44.8 (44.9) for the neurology and MRI group, respectively. This trend approached statistical significance (p=0.075).

The number of headache days and mean pain score in the three months prior to recruitment were assessed at baseline using the MIDAS questionnaire (Table 36). The mean (SD) number of headache days was estimated at 51.6 (31.5) and 42.8 (30.7) for the neurology and MRI groups, respectively, with the difference being statistically significant (p=0.038). The mean (SD) headache pain was estimated at 6.9 (1.8) and 6.9 (1.9) for the neurology and MRI groups (p=0.778).

The disability score at baseline as per the MIDAS questionnaire is presented in Table 37. A higher proportion of patients in the neurology group (70%) presented a grade IV (severe disability) compared to the MRI group (60%). However, no statistically significant difference in the distribution of grade of headache disability between both groups (p=0.357).

Table 37. MIDAS grade disability score at baseline.

		Neurology		MRI		p-value *
		N	%	N	%	
MIDAS Grade	Little or No Disability	11	8.6%	13	14%	p=0.357
	Mild Disability	12	9.4%	7	7.4%	
	Moderate Disability	11	8.6%	13	14%	
	Severe Disability	90	70 %	57	60%	
	Missing	4	3.1%	5	5.3%	

Notes: * Chi-square test.

NHS healthcare resources utilised in the management of chronic headache in the 12 months prior to study recruitment were collected, including GP appointments and all NHS appointments (including GP appointments), are presented in Table 38. The neurology group had a statistically significant higher number of both GP appointments [mean (SD) of 3.7 (2.9) vs 2.4 (1.5), p<0.001] and all NHS appointments [mean of 4.3 (3.7) vs 2.5 (1.5), p<0.001].

Table 39 summarises the respective costs in the 12 months prior to recruitment, showing that the participants in the neurology group had higher costs compared to the MRI group participants [mean (SD) cost per participant of £231 (£294) vs £106 (£121), GLM analysis p<0.001].

Table 38. GP and NHS contacts associated with the management of chronic headache in the 12 months prior recruitment to the study.

	Neurology group (n=128)		MRI group (n=95)		p-value
Type of NHS appointment	**Total of episodes**	**Mean (SD)**	**Total of episodes**	**Mean (SD)**	
GP appointments	474	3.7 (2.9)	220	2.4 (1.5)	<.001
All NHS appointments (i.e. primary and secondary care appointments)	552	4.3 (3.7)	235	2.5 (1.5)	<.001

Table 39. GLM analysis for the 12-month cost prior recruitment variable (values in £, 2018).

	Neurology group (n=128)	MRI group (n=95)	Difference (MRI-Neurology)	95% CI	p-value
Total Cost during the 12 months prior to recruitment	231 (294)	106 (121)	-125	-181 to -70	< 0.001

4.3.4 Primary Outcome

The primary outcome was to estimate the 6-month costs condition-related associated with both groups. Table 40 summarises the 6-month cost distribution (mean, standard deviation, minimum, median, percentiles 25 and 75 and maximum) per group. The mean cost of management per participant [mean (SD)] was lower in the MRI group compared to the neurology group [£245 (£172) vs £578 (£420)] (Table 40), leading to a mean cost difference between groups of £333 per participant.

Table 40. Descriptive statistics of six month costs associated with the neurology and the MRI group.

	Mean	Standard Deviation	Minimum	Percentile 25	Median	Percentile 75	Maximum
Neurology (n=128)	£578	£420	£40	£396	£476	£628	£3,758
MRI (n=95)	£245	£172	£44	£158	£188	£263	£1,494

The cost distribution is positively skewed (mean >> median), as it is affected by a small proportion of patients that had significantly higher costs. A higher proportion of patients in the MRI group were in the £0-£250 range compared to the Neurology group (71.6% vs 4.7%). In the over £1000 category, a lower proportion of participants in the MRI group were found compared to the control group (1.1% vs 7.0%, with a maximum cost of £3,758 and £1,494 for the neurology and MRI group, respectively). Eighty-five percent of participants in the control group cost between £250 and £750 compared to 27% in the MRI group. The latter is illustrated in Figure 53 and Figure 54.

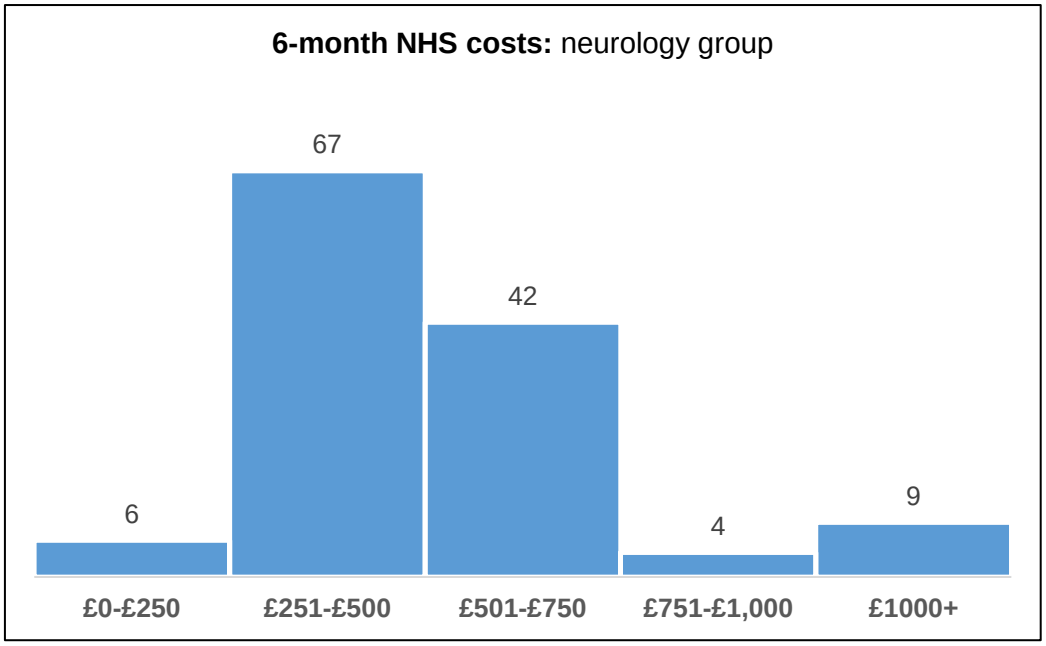


Figure 53. Histogram for the 6-month cost distribution for the neurology group.

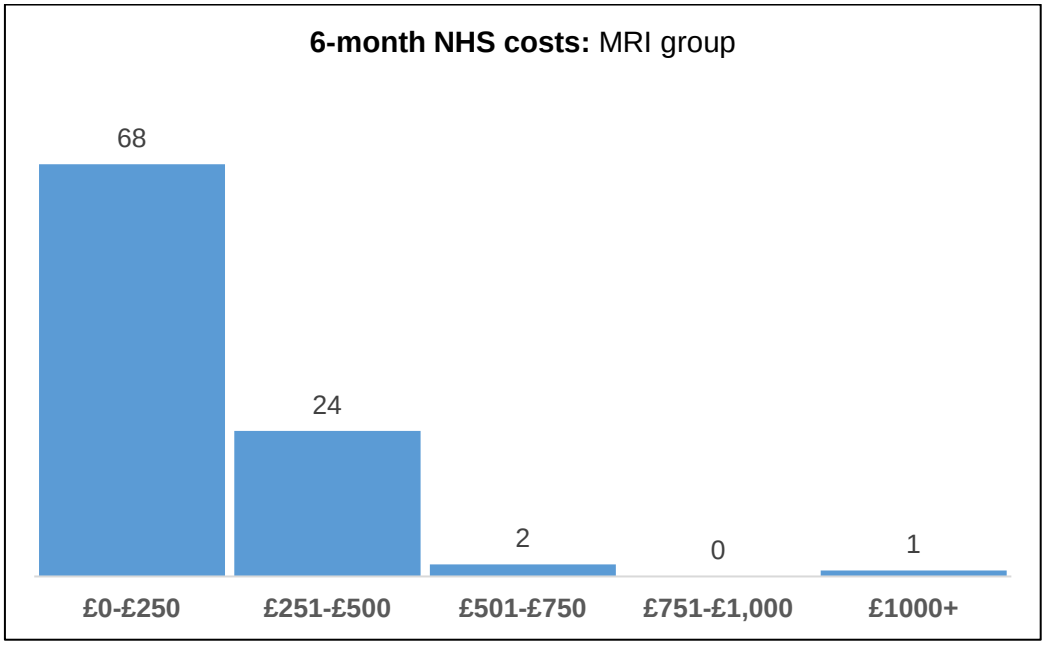


Figure 54. Histogram for the 6-month cost distribution for the MRI group.

(i) GLM analysis: ***Unadjusted Generalised Linear Model***

Taking into consideration the anticipated data skewness, the statistical analysis plan used GLM to model the 6-month cost analysis, using a Gamma family distribution and identity link. The analysis is summarised in Table 41.

Table 41. GLM analysis for the 6-month cost analysis variable (values in £, 2018).

	Neurology group (n=128)	**MRI group (n=95)**	**Difference (MRI-Neurology)**	**95% CI**	**p-value**
Total cost at 6 months post-recruitment	578 (420)	245 (172)	-333	-413 to -253	< 0.001

The mean cost difference between both groups at 6 months was -£333 (CI 95%: - £413 to -£253) per participant (p<0.001).

(ii) GLM analyses: ***Adjusted Generalised Linear Model***

Given the non-randomised study design, a second GLM analysis was performed to adjust using baseline characteristics that were statistically different between groups (see Table 42). A conservative threshold (p<0.10) was assumed for this purpose. The variable 'ethnicity' was not included as it was a string variable. For instance, female patients (variable 'gender') had costs that were on average £53 higher than men (see Table 42). The mean cost difference between both groups at 6 months after adjusting for baseline imbalances decreased to -£308 (CI 95%: -£408 to -£209) per participant (p<0.001).

Table 42. GLM analysis for the 6-month cost analysis variable (gamma function, identity link) adjusted using statistically significant differences (p<0.1) between groups at baseline.

6-month cost	**Coef.**	**Std. Err.**	**t**	**P>|t|**	**[95% CI]**	
Group (Neurology/MRI)	-308.3	50.8	-6.07	0.000	-407.8	-208.8
Gender	53.1	37.2	1.43	0.154	-19.9	126.1
Time referral MRI or Neurologist	0.018	0.305	0.06	0.954	-0.580	0.615
Baseline EQ-5D-5L utility	0.318	1.155	0.27	0.783	-1.946	2.581
Baseline EQ-5D-5L VAS	-0.341	1.170	-0.29	0.771	-2.634	1.952
Baseline HIT-6	0.020	0.123	0.17	0.868	-0.220	0.261
Baseline MIDAS score	0.121	0.463	0.26	0.793	-0.785	1.028
Baseline MIDAS Headache Days	-0.164	0.468	-0.35	0.727	-1.082	0.755
NHS costs 12 months before	0.139	0.087	1.59	0.112	-0.032	0.310
Constant	780.1	141.3	5.52	0.000	503.2	1057.1

Besides the study's observational nature, the potential impact on the primary outcome due to missing follow-up data was also assessed. For this purpose, differences in baseline characteristics between participants with complete follow-up up to 6 months compared to those without follow-up were assessed. A third GLM, including the follow-up variable (binomial Yes/No), was performed and the results are presented in Table 43. The mean cost difference between both groups at 6 months after adjusting for baseline imbalances decreased to -£302 (CI 95%: -£411 to -£193) per participant (p<0.001)

Table 43. GLM analysis for the 6-month cost analysis variable (gamma function, identity link) adjusted using statistically significant differences (p<0.1) between groups at baseline and considering the presence/absence of complete follow-up up to month 6.

6-month cost	**Coef.**	**Std. Err.**	**t**	**P>|t|**	**[95% CI]**	
Group (NE/RA)	-302.3	55.7	-5.430	0.000	-411.4	-193.1
Gender	53.0	37.4	1.410	0.157	-20.4	126.3
Time from referral to MRI or appointment with a neurologist	0.024	0.304	0.080	0.938	-0.573	0.620
Baseline EQ-5D-5L utility	0.390	1.188	0.330	0.743	-1.938	2.718
Baseline EQ-5D-5L VAS	-0.414	1.203	-0.340	0.731	-2.773	1.945
Baseline HIT-6	0.017	0.124	0.140	0.890	-0.227	0.261
Baseline MIDAS score	0.111	0.461	0.240	0.809	-0.792	1.015
Baseline MIDAS Headache Days	-0.151	0.468	-0.320	0.747	-1.067	0.766
NHS costs 12 months before	0.134	0.088	1.530	0.127	-0.038	0.306
Follow-up	13.1	49.6	0.260	0.792	-84.1	110.3
Constant	772.2	145.1	5.320	0.000	487.8	1056.6

Table 44 summarises the three GLM analyses performed. The initial unadjusted 6-month cost difference between groups (-£333) was robust and not too affected by the observational nature of the study (-£308) or follow-up completeness (-£302). In all analyses, direct access to MRI for the management of chronic headache was associated with statistically significant NHS cost savings.

Table 44. Summary table with the three GLM analysis for the 6-month cost analysis: (i) unadjusted; (ii) adjusted for imbalance in baseline characteristics; and (iii) adjusted for follow-up completeness.

	Mean 6-month cost [95% CI]		
	Unadjusted	Adjusted for baseline characteristics (p<0.1)	Adjusted for baseline characteristics and follow-up completeness (p<0.1)
Primary Outcome (6-month cost analysis)	-£333 [-£413, -£253]	-£308 [-£408, -£209]	-£302 [-£411, -£193]

(ii) Bootstrap analysis

A 1000-replicate bootstrap analysis for the variable total 6 month costs was also performed. The 95% confidence interval for three types of bootstraps (normal, percentile and bias-corrected) are presented in Table 45. As with the GLM, the same mean difference cost per participant was obtained (- £333) but the 95% confidence intervals varied based on the type of bootstrap analysis considered: normal (CI 95%: - £413 to -£253); percentile (CI 95%: - £414 to -£257); and bias-corrected (CI 95%: - £419 to -£263). All bootstrap analyses showed a statistically significant cost difference per participant (the value 0 is not included in the 95% CI). The use of bias-corrected bootstrap analysis is the most appropriate to assess skewed cost data (Jiang and Zhou 2004).

Table 45. Bootstrap analysis for the variable total cost at 6 months (1,000 replicates).

Variable	Reps	Observed	Bias	Std. Err.	[95% Conf. Interval]		
Study Group	1000	-£333	1.132	40.7	-£413	-£253	(N)
					-£414	-£257	(P)
					-£419	-£263	(BC)

Note: N=normal; P=percentile; BC=bias-corrected

4.3.5 Secondary Outcome

1. *To estimate the 12-month costs following the initial episode at secondary care associated with direct access to MRI compared with referral to the neurology department.*

The first secondary outcome extended the 6 month cost analysis to the 12 months following recruitment into the study. Table 46 summarises the total and mean number of NHS events up to 12 months post-recruitment, along with the number and percentage of participants responsible for such events. With regards to primary care appointments, participants in the neurology group had a higher number of GP face-to-face appointments (mean of 1.82 vs 1.19, p=0.006). If secondary care appointments are considered, participants in the neurology group had a higher number of outpatient appointments (mean of 2.52 vs 0.26, p<0.001) and other treatments such as Botox and nerve injection (mean of 0.30 vs 0.05, p<0.001). In contrast, participants in the neurology group had a lower number of head MRIs (mean number 0.59 vs 1.05, p<0.001). There were no statistically significant differences in utilisation of the remaining healthcare episodes.

Table 46. Number of NHS appointments organised per type of activity and study group.

	Neurology group (n=128)			MRI group (n=95)			p-value
Type of NHS appointment	**Total episodes**	**Mean (SD)**	**N (%)**	**Total episodes**	**Mean (SD)**	**N (%)**	
Primary care services							
GP face-to-face appointment	233	1.82 (2.11)	128 (100)	113	1.19 (1.64)	82 (86)	0.006
GP phone appointment	37	0.29 (0.75)	24 (19)	25	0.26 (0.49)	23 (24)	0.420
Hospital based services							
Hospital outpatient appointment	322	2.52 (1.19)	128 (100)	25	0.26 (0.55)	20 (21)	<0.001
Inpatient episode	4	0.03 (0.35)	1 (0.8)	1	0.01 (0.10)	1 (1.1)	0.837
ED episode	8	0.06 (0.24)	8 (6.3)	5	0.05 (0.22)	5 (5.3)	0.756
Head CT	1	0.01 (0.09)	1 (0.8)	1	0.01 (0.10)	1 (1.1)	0.832
Head MRI	75	0.59 (0.49)	75 (59)	100	1.05 (0.30)	95 (100)	<0.001
Others (e.g. Botox and nerve injection treatments)	39	0.30 (0.79)	25 (20)	5	0.05 (0.30)	3 (3.2)	<0.001

The use of NHS resource pre and post-recruitment were also compared. Table 47 summarises the difference between headache-related events 12 months post-recruitment compared to the 12 months pre-recruitment. A reduction of 197 and 87 GP visits, equivalent to a mean reduction of 1.54/0.92 per participant, was reported in the neurology and MRI group, respectively. Similarly, there was a reduction in emergency department utilisation with a decrease of 23 and 6 episodes, equivalent to a mean reduction of 0.18 and 0.07 episodes per participant, in the neurology and MRI group, respectively.

Table 47. Differences in the total number, mean and percentage reduction of NHS appointments organised per type of activity and study group in the 12 months post-recruitment compared to the 12 months pre-recruitment.

	Neurology group (n=128)		MRI group (n=95)	
Type of NHS appointment	**Total episodes**	**Mean**	**Total episodes**	**Mean**
Primary care services				
GP face-to-face appointment	-197	-1.54	-87	-0.92
GP phone appointment	-11	-0.09	5	0.05
Hospital-based services				
Hospital outpatient appointment	301	2.36	23	0.24
Inpatient episode	3	0.02	1	0.01
Emergency Department episode	-23	-0.18	-6	-0.07
Head CT	-9	-0.07	1	0.01
Head MRI	55	0.43	100	1.05

Note: a negative/positive number denotes a decrease/increase in activity following recruitment.

(i) GLM analysis: ***Unadjusted Generalised Linear Model***

Similarly to the primary outcome, a GLM analysis using the study group as the only covariate was conducted to estimate the 12-month total costs (Table 48). As with the 6-month cost analysis, the MRI intervention was associated with lower overall cost, with a mean cost difference between groups of -£518 (CI 95%: - £637 to - £401) per participant (p<0.001).

Table 48. GLM analysis for 12-month costs (cost values expressed in £).

	Neurology group (n=128)	MRI group (n=95)	Difference (MRI-Neurology)	95% CI	P value
Total Cost at 12 months	808 (579)	289 (246)	-518	-637 to -401	< 0.001

(i) GLM analyses: ***Adjusted Generalised Linear Model***

Two further adjusted GLM analyses were performed and the results are presented in Table 49. As with the primary outcome, the intervention with direct access to MRI for the management of chronic headache was associated with cost savings despite the observational nature of the study and the absence of follow-up data for some participants.

Table 49. Summary table with the three GLM analysis for the 12-month cost analysis: (i) unadjusted; (ii) adjusted for imbalance in baseline characteristics due to non-randomised design; and (iii) the addition of differences in terms of follow-up completeness.

	Mean 12-month cost [95% CI]		
	Unadjusted	Adjusted for baseline characteristics (p<0.1)	Adjusted for baseline characteristics follow-up completeness (p<0.1)
Secondary Outcome (12-month cost analysis)	-£518 [-£637, -£401]	-£496 [-£639, -£353]	-£481 [-£635, -£327]

(ii) Bootstrap analysis

A 1000-replicate bootstrap analysis for the variable total 12 month costs was also performed. The 95% confidence intervals for three types of bootstraps (normal, percentile and bias-corrected) are presented in Table 50. All bootstrap analyses showed a statistically significant cost difference per participant (the value 0 is not included in the 95% CI).

Table 50. Bootstrap analysis for the variable total cost at 6 months (1,000 replicates).

Variable	Reps	Observed	Bias	Std. Err.	[95% Conf. Interval]		
Study Group	1000	-£518	1.188	57.5	-£632	-£406	(N)
					-£632	-£410	(P)
					-£634	-£411	(BC)

Note: N= normal; P= percentile; BC= bias-corrected

2. *To perform cost-effectiveness analyses at 6 and 12-month following direct access to MRI compared with referral to the neurology department.*

Other secondary outcomes related to cost-effectiveness analyses using measures of effect based on a self-reported non-headache specific questionnaire (EQ-5D-5L questionnaire). This is a cost-utility analysis. The utility values and health rating (based on a 0-100 visual analogue scale) at baseline, month 6 and 12 are presented in Table 51.

At baseline, participants in the neurology group showed a trend of lower utility (mean utility of 0.809 vs 0.830, p=0.097) and a statistically significant lower self-perceived health rating (mean score of 64.0 vs 70.8, p=0.005). No statistically significant difference (p>0.05) between the groups occurred at either 6 or 12 months in relation to the utility and self-reported VAS scores.

A high proportion of data were missing, particularly at 12 months, with 75% and 85% of follow-up data missing in the neurology and MRI groups, respectively. For this reason, it was decided not to undertake a cost-utility analysis at 12 months. Hence, this section presents the 6-month cost-utility analysis only.

Table 51. Descriptive statistics for the variable utility and VAS values.

			N	Mean	Standard Deviation	p-value
BASELINE	Utility	Neurology	127	0.809	.182	0.097
		Radiology	95	0.830	.195	
	VAS	Neurology	126	64.0	18.8	0.005
		Radiology	95	70.8	20.2	
MONTH 6	Utility	Neurology	55	0.770	.263	0.243
		Radiology	26	0.681	.346	
	VAS	Neurology	53	68.2	20.6	0.463
		Radiology	23	62.5	24.7	
MONTH 12	Utility	Neurology	32	0.667	.335	0.548
		Radiology	15	0.631	.369	
	VAS	Neurology	28	65.5	21.7	0.533
		Radiology	15	55.5	30.3	

Cost-effectiveness analyses were based on incremental costs divided by incremental effects, in this case measured in QALYs. The mean cost per QALY at month 6 (Equation 9) was estimated at £7,483.

Equation 9. Estimate of the incremental cost per QALY at month 6.

$$ICER = \frac{Cost\ MRI - Cost\ neurology}{QALY\ MRI - QALYneurology} = \frac{£245 - £578}{0.341 - 0.385} = \frac{-£333}{-0.045} = £7{,}483$$

In order to adjust for imbalance in mean baseline utilities, a multiple regression analysis was used as suggested by Manca, Hawkins, and Sculpher (2005) and detailed below in Table 52 and Equation 10.

Table 52. Summary of the regression analysis for utility at month 6 adjusted by two variables: randomisation group and baseline utility.

Number of observations=76; R-squared = 0.2440; Adjusted R-squared = 0.2233

Utility month 6	Coef.	Std. Err.	t	P>|t|	[95% Conf. Interval]	
Group	-0.1206	0.0655	-1.84	0.07	-0.2512	0.0010
Utility baseline	0.8994	0.1932	4.66	0.00	0.5144	1.2843
Constant	0.1495	0.1762	0.85	0.399	-0.2017	0.5007

Equation 10. Multiple regression analysis for the utility at month 6 adjusted per study group and utility at baseline.

*Utility at month 6 = 0.1495 - 0.1206 * Study Group + 0.8994 * Utility at baseline*

Note: Study group (0= neurology group; 1= MRI group).

Figure 55 illustrates the bootstrap analysis with 1,000 replicates, considering the 6-month cost per QALY adjusted using the baseline utility. At month 6, the intervention with MRI had a probability of 4.1% of being dominant (i.e. increased QALYs at a lower cost) compared to the neurology group. In the obverse quadrant, there was a 0.0% probability of the MRI intervention being dominated (i.e. lower QALYs at a higher cost) by the neurology group. The remaining 95.9% of bootstraps were in the cost-effectiveness analysis quadrants, i.e. the probability of being cost-effective depended on the overall system willingness-to-pay for each QALY.

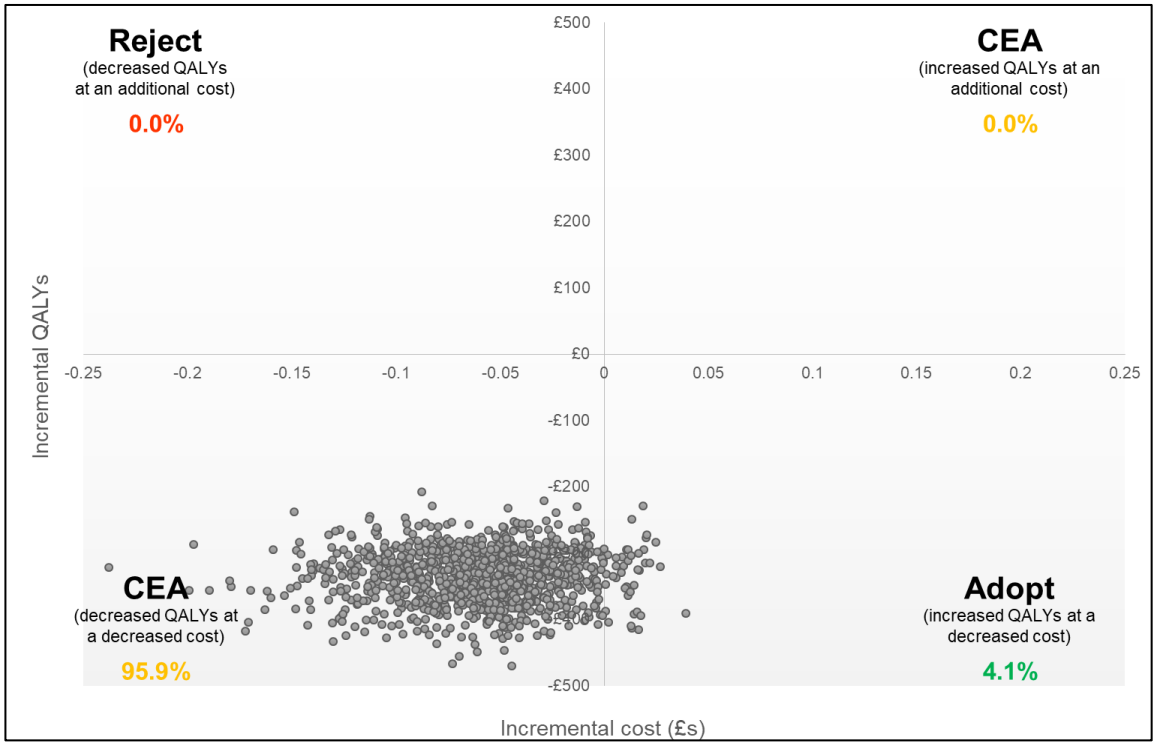


Figure 55. Cost-effectiveness plane associated with the 6-month cost per QALY analysis and probability associated each quadrant (bootstrap analysis with 1,000 replicates).

The cost-effectiveness acceptability curve (CEAC) shows the probability of access to MRI being more cost-effective than referral to neurology at different willingness-to-pay thresholds (Figure 56). For instance, assuming a £20,000 and £30,000 willingness-to-pay per QALY (thresholds typically considered by NICE), there was a 7.2% and 4.1% probability that direct access to MRI was more cost-effective at month 6, respectively (Figure 56). Unlike the usual situation with replicates located in the northeast quadrant (more costly and more effective), in this case the higher the system's willingness-to-pay threshold, the less likely the intervention is to be cost-effective. This is because most replicates were within the southwest quadrant (less costly and less effective). The CEAC cuts the y axis at 1.0 as it is the probability of the MRI group generating cost savings and asymptotes towards 0 as four percent of replicates generated health gains (Fenwick, O'Brien, and Briggs 2004).

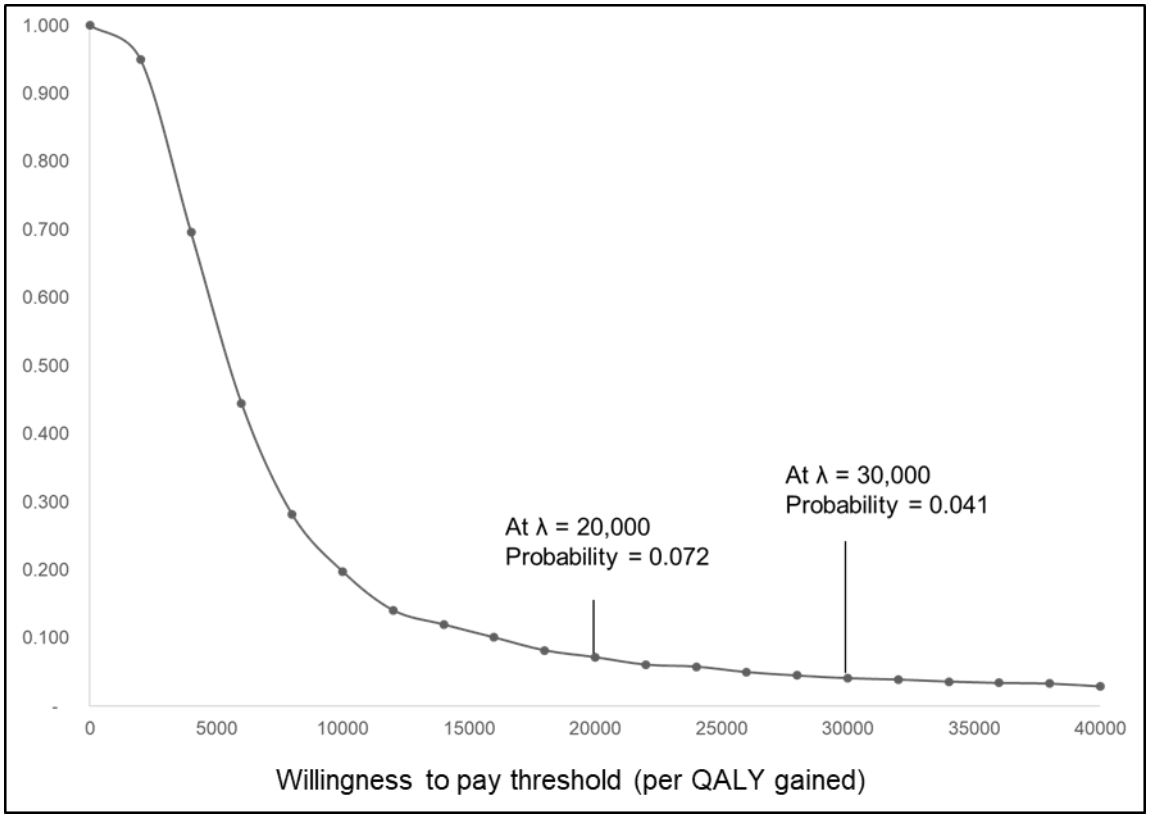


Figure 56. CEAC for several thresholds of willingness-for-pay (represented as λ).

3. *To compare the levels of patient satisfaction associated with the two pathways.*

Participant satisfaction was evaluated at month 3 post-recruitment using a non-validated questionnaire. This questionnaire evaluated three dimensions: (a) referral process (time elapsed between referral from primary care to initial secondary care appointment) (Table 53); (b) initial appointment (on the day satisfaction) (Table 54); and (c) the overall experience three months after recruitment (Table 55). The Pearson Chi-square statistical test was used in all analyses.

The first two questions assessed the period elapsed between the GP referral and the initial secondary appointment (Table 53), either for a neurologist appointment or a head MRI. More participants in the MRI group tended to receive an appointment in an acceptable timeframe compared to those in the neurology group (p=0.193). In addition, participants in the MRI group

reported higher satisfaction levels (p=0.005) associated with the information received prior to the actual appointment.

The next two questions evaluated the initial appointment itself (Table 54). No statistical difference (p=0.366) between the groups was found regarding the satisfaction levels with both types of appointment (MRI scan or neurologist initial appointment). A higher proportion of participants in the neurology group reported a better experience compared to their expectation (p=0.002).

The final six questions (Table 55) assessed the overall patient experience at month 3 post-recruitment. In all variables, except frequency of appointments (p=0.166), participants in the neurology group reported higher levels of satisfaction with: amount of time spent with clinical staff (p=0.001); consistency of care (p=0.028); how informed you felt about your condition (p=0.010); how informed you felt about your treatment (p=0.004); and the overall experience (p<0.001).

Table 53. Patient experience questionnaire associated with the referral process to either neurology (n=99) or MRI group (n=79).

Before the appointment		Not sure		Yes		No	
		N	%	N	%	N	%
Did you receive your appointment within a timeframe acceptable to you?	Neurology (n=99)	8	8.1%	66	67%	25	25%
	MRI (n=79)	3	3.8%	62	79%	14	18%

		Very satisfied		Satisfied		Neutral		Dissatisfied		Very dissatisfied	
		N	%	N	%	N	%	N	%	N	%
How satisfied were with the information you received beforehand?	Neurology (n=98)	25	26%	48	49%	22	22%	3	3.1 %	0	0%
	MRI (n=79)	38	50%	24	32%	10	13%	2	2.6 %	2	2.6 %

Table 54. Patient experience questionnaire associated with either the Neurology (n=99) or the MRI group (n=50) appointment.

Regarding the appointment		Very satisfied		Satisfied		Neutral		Dissatisfied		Very dissatisfied	
		N	%	N	%	N	%	N	%	N	%
How you found the process overall	Neurology (n=98)	55	56%	36	36%	5	5.1 %	3	3.0 %	0	0%
	MRI (n=50)	27	54%	22	44%	0	0.0 %	1	2.0 %	0	0%

		A better experience		About the same as expected		A worse experience	
		N	%	N	%	N	%
How did you find the experience in comparison to what you had expected?	Neurology (n=95)	57	60%	35	37%	3	3.2%
	MRI (n=45)	13	29%	31	69%	1	2.2%

Table 55. Overall patient experience for participants in the Neurology (n=56) or the MRI group (n=14) group.

About the entire experience at 3 months post-recruitment		Very satisfied		Satisfied		Neutral		Dissatisfied		Very dissatisfied	
		N	%	N	%	N	%	N	%	N	%
Amount of time spent with clinical staff	Neurology (n=56)	20	36%	30	54%	6	11%	0	0%	0	0%
	MRI (n=14)	3	21%	3	21%	5	36%	1	7.1%	2	14%
Consistency of care	Neurology (n=29)	16	55%	0	0%	12	41%	1	3.4%	0	0%
	MRI (n=14)	2	14%	3	21%	6	43%	2	14.3%	1	7.1%
Frequency of appointments	Neurology (n=55)	9	16%	19	35%	22	40%	4	7.3%	1	1.8%
	MRI (n=14)	2	14%	1	7.1%	7	50%	3	21%	1	7.1%
How informed you felt about your condition	Neurology (n=54)	15	28%	24	44%	9	17%	5	9.3%	1	1.9%
	MRI (n=13)	2	15%	2	15%	5	39%	1	7.7%	3	23%
How informed you felt about your treatment	Neurology (n=55)	15	27%	18	33%	17	31%	4	7.3%	1	1.8%
	MRI (n=14)	1	7.1%	1	7.1%	5	36%	5	36%	2	14%
Your overall experience	Neurology (n=55)	13	24%	32	58%	7	13%	3	5.5%	0	0%
	MRI (n=14)	2	14%	1	7.1%	9	64%	1	7.1%	1	7.1%

4. *To estimate the disease burden (using headache diaries and HIT-6 and MIDAS questionnaires) associated with both pathways.*

Two headache-specific questionnaires were used to assess the headache burden associated with both groups (HIT-6 and MIDAS questionnaire).

The Headache Impact Test (HIT-6) questionnaire (Kosinski et al. 2003) measured the headache burden based on 6 questions, leading to a score range from 36 to 78. The higher the score, the more severe the headache burden is.

Participants in the neurology group reported a higher headache burden at baseline compared to the MRI group (mean score of 65.0 vs 62.6, p=0.006) (Table 56). Over the follow-up period no statistically significant differences between the groups, both at 6 months (mean HIT-6: 60.0 vs 53.1, p=0.968) and 12 months (62.0 vs 56.5, p=0.409). The use of the HIT-6 questionnaire was associated with very high attrition rate (79% and 88% at month 6 and 12, respectively).

Table 56. Descriptive statistics for the HIT-6 questionnaire.

			N	Mean	Standard Deviation	p-value
BASELINE	HIT-6 Score	Neurology	128	65.0	5.3	0.006
		MRI	92	62.6	7.3	
MONTH 6	HIT-6 Score	Neurology	35	60.0	8.7	0.968
		MRI	12	53.1	22.8	
MONTH 12	HIT-6 Score	Neurology	21	62.0	9.9	0.409
		MRI	5	56.5	10.3	

The Migraine Disability Assessment (MIDAS) questionnaire (Stewart et al. 2001) also assessed headache disability. The higher the score (ranges from 0 to 90), the higher the MIDAS grade and the respective headache severity. Participants in the neurology group presented a trend of higher headache severity at baseline compared to the MRI group (mean MIDAS score of 57.8 vs 44.8, p=0.075) (Table 57). Over the follow-up period, neurology participants reported more severe headaches both at 6 months (52.6 vs 40.7, p=0.827) and 12 months (45.7 vs 29.3, p=0.498) but both differences were not statistically significant. Similar to the HIT-6 questionnaire, the use of MIDAS questionnaire was associated with very high attrition rate (80% and 82% at month 6 and 12, respectively).

A second variable estimated the number of headache days. Participants in the neurology group reported a higher number of headache days at baseline compared to the MRI group (51.6 vs 42.8, p=0.038). A similar trend seemed to be kept over the follow-up period, both at 6 months (41.2 vs 26.5, p=0.152) and 12 months (36.1 vs 9.9, p=0.170).

A third variable assessed self-reported headache pain scores. Participants at baseline showed similar headache pain scores (6.9 vs 6.9, p=0.827). During the follow-up period, no statistical

differences were found at 6 months (4.5 vs 4.2, p=0.663) or 12 months (5.5 vs 6.8, p=0.407). As with other outcomes, very high attrition rates were reported over time, particularly at 12 months post-recruitment, and in both groups, mostly in the MRI group.

Table 57. Descriptive statistics for the MIDAS questionnaire.

			N	Mean	Standard Deviation	p-value
BASELINE	MIDAS Score	Neurology	124	57.8	54.0	0.075
		MRI	90	44.8	44.9	
	Headache days	Neurology	124	51.6	31.5	0.038
		MRI	90	42.8	30.7	
	Pain Score	Neurology	124	6.9	1.8	0.778
		MRI	90	6.9	1.9	
MONTH 6	MIDAS Score	Neurology	32	52.6	58.1	0.827
		MRI	12	40.7	36.0	
	Headache days	Neurology	46	41.2	28.6	0.152
		MRI	15	26.5	28.1	
	Pain Score	Neurology	44	4.5	2.3	0.663
		MRI	15	4.2	2.0	
MONTH 12	MIDAS Score	Neurology	29	45.7	55.7	0.498
		MRI	10	29.3	37.0	
	Headache days	Neurology	30	36.1	35.8	0.170
		MRI	10	9.9	9.7	
	Pain Score	Neurology	29	5.5	2.5	0.407
		MRI	10	6.8	3.0	

Changes in management:

Changes to care management in both groups were evaluated. It was considered that a change in care management occurred when patients were prescribed new medication or underwent new treatments, e.g. nerve block injection or Botox treatments.

Participants in the neurology group were more likely to have a change in therapeutic management compared to participants in the MRI group (97% vs 64%, p<0.001). Similarly, among participants that had not started on preventative medication pre-recruitment, a higher proportion of participants in the neurology group were started on preventative medication as part of their clinical management (93% vs 53%, p<0.001).

Incidental findings:

Out of the 95 participants recruited into the MRI group, 3 MRIs were not performed during the initial appointment due to claustrophobia events. From a total number of 92 MRIs, 85 (92%) were normal and 7 (7.6%) presented abnormal findings (Table 58), with one diagnosis being particularly significant (two small intracranial aneurysms). This participant was referred to the neurovascular team for assessment, at which point no intervention was performed during the follow-up period (participant included in an aneurysm active surveillance group). Two other participants had a change in their clinical management for less significant findings, leading to either a follow-up appointment or an imaging scan. No brain malignancies were diagnosed.

Table 58. Description of incidental findings, clinical relevance and subsequent pathway.

Abnormal findings	**Significant (Yes/No)?**	**Changes in diagnostic or treatment pathway**
Mature striatocapsular lacune	No	
Sinusitis with complete opacification	No	Ear, nose and throat (ENT) review only.
Pituitary abnormality	No	
Low lying cerebellar tonsils	No	
Previous petrous surgery noted	No	
Two aneurysms: anterior communicating artery and right internal carotid artery	Yes	Referred to neurovascular – no coiling (no intervention).
No definitive lesion	No	Follow-up MRI only.

5. *To evaluate time-off work due to the chronic headache associated with the pathway with referral to the neurology department compared with direct access to MRI.*

The number of days off work were estimated using participant diaries and MIDAS questionnaires. Participants in the neurology group (n=83) reported higher mean number of days off work due to headache compared to the MRI group (n=35) but these were not statistically significant at 6 months (13.9 vs 9.7, p=0.563) or 12 months (27.9 vs 19.1, p=0.808) post-recruitment (Table 59).

Table 59. Descriptive statistics of the number of days off work per group (n=83 for the neurology group and n=35 for the MRI group).

			Mean	Standard Deviation	p-value
0-6 Months	Days off work	Neurology	14.0	22.7	0.883
		MRI	9.3	12.1	
6-12 Months	Days off work	Neurology	13.9	22.6	0.563
		MRI	9.7	7.5	
0-12 Months	Days off work	Neurology	27.9	44.7	0.809
		MRI	19.1	18.7	

4.3.6 Sensitivity Analyses

Deterministic sensitivity analyses around three unit cost parameters were performed: (i) head MRI; (ii) neurology appointments (both first and follow-up outpatient appointments); and (iii) 'Did not attend' cost. One additional sensitivity scenario considered a societal perspective of analysis instead of the NHS perspective. The results from the sensitivity analyses are summarised in Table 60.

(i) Direct access to MRI. The unit cost per head MRI was a fundamental cost variable associated with participants in the MRI group. The £146 unit cost was based on the NHS reference costs, code "RD01A (imaging: direct access): Magnetic Resonance Imaging Scan of One Area, without Contrast, 19 years and over" (NHS Improvement 2017). This unit cost considered not only to the time it took to complete the actual acquisition time but also the subsequent reporting. This unit cost was varied - 25%, equivalent to £110. This change was consistent with the historical trend of reduction exhibited in the direct access to MRI tariff. Under this scenario, the cost difference per participant at 6 and 12 months between groups increased, respectively, to -£367 ($p<0.001$) and -£552 ($p<0.001$).

(ii) Neurology appointments. The unit cost per neurology appointment (first and follow-up appointments) was reduced by 25%, equivalent to £180 and £111 for the initial and follow-up neurology appointment, respectively. Again, this variation was consistent with the historical trend of decrease in the unit cost per neurology consultation in the NHS reference costs. This unit cost variation led to a decrease of the cost difference between groups at 6 and 12 months to, respectively, -£263 ($p<0.001$) and -£482 ($p<0.001$).

(iii) 'Did not attend' (DNA) assumption. The baseline unit cost per DNA event was considered at 30% of the cost of the event when patients did attend. The latter is particularly relevant in outpatient visits (i.e. neurology appointments). This deterministic sensitivity analysis did not consider DNA costs (i.e. equals to £0). This variation led to a mean cost difference between groups of -£327 ($p<0.001$) at month 6 and -£499 ($p<0.001$) at month 12.

(iv) Societal costs. In July 2019, the average full-time weekly earnings was estimated at £503 (Office for National Statistics 2019a), equivalent to a daily wage of £100.6 per work day. The daily wage of participants in part-time occupation or participants unemployed but actively looking for a job were assumed to be half of those in full-time occupation. Participants permanently sick or disabled, in education (college or university) or wholly retired from work were assumed to have no productivity loss (£0).

Table 60. Sensitivity analyses scenarios considered and respective impact on the cost analyses at month 6 and 12.

	Mean cost difference		**P-value (using GLM)**	
	Month 6	Month 12	Month 6	Month 12
Base case scenario	**-£333**	**-£518**	**<0.001**	**<0.001**
(i) **Head MRI:** **-25%** of unit cost of the head MRI	-£367	-£552	<0.001	<0.001
(ii) **Neurology appointments:** **-25%** of unit cost of the initial/follow-up neurology outpatient appointments	-£263	-£482	<0.001	<0.001
(iii) **'Did not attend' cost:** No DNA cost (baseline=30%).	-£327	-£499	<0.001	<0.001
(iv) **Societal costs.** The use of a **societal perspective of analysis**.	-£417	-£769	0.086	0.037

Table 61 summarises the findings from non-adjusted and adjusted GLM cost analyses incorporating both NHS and time off work related costs from Table 59. There was a trend of cost savings up to month 6 associated with the use of direct access to head MRI in both GLM analyses. When these analyses were extended up to 12 months, the cost difference between groups was statistical significant ($p<0.05$) for both the non-adjusted and adjusted GLM analyses.

Table 61. Summary table with the two GLM analysis for the 6 and 12-month total cost analyses: (i) unadjusted; (ii) adjusted for imbalance in baseline ($p<0.1$).

	Societal perspective [95% CI, p-value]	
	Unadjusted	Adjusted for baseline characteristics ($p<0.1$)
Mean total costs 0-6 months	-£417 [-£1,240, £406] p=0.321	-£576 [-£1234 , £81] p=0.086
Mean total costs 0-12 months	-£769 [-£1,452, -£85] p=0.027	-£896 [-£1,739, -£54] p=0.037

4.3.7 Summary of Results

Direct access to head MRI from primary care led to lower mean cost per participant at 6 and 12 months compared to the conventional route with referral for a neurology appointment. Based on adjusted GLM distributions, the mean cost difference per participant between groups at 6 and 12-month post recruitment were estimated at, respectively, -£302 (95% CI: -£411, -£193) and -£481 (95% CI: -£635, -£327). When productivity losses were considered, the mean cost difference between groups increased as participants in the neurology group were more likely to report time off work and loss of productivity.

However, using the £20,000 and £30,000 willingness-to-pay per QALY thresholds (typically considered by NICE), there was only a 7.2% and 4.1% probability of direct access to MRI being more cost-effective than neurology referrals at month 6, respectively. The MRI pathway delivered statistically lower costs than the neurology pathway, but poorer health outcomes measured in QALYs and was not cost-effective using usual NICE willingness-to-pay thresholds.

At baseline, participants in the neurology group self-reported higher headache burden and lower health scores. These findings were not sustained over time at month 6 or 12 post-recruitment.

Participant satisfaction analyses showed that participants in the MRI group presented higher satisfaction levels during the referral process, particularly associated with the information provided ($p<0.05$). However, on the day of the first appointment (either a neurology appointment or head MRI scan), participants in the neurology group reported a better experience compared to their expectation. In addition, 3 months post-recruitment, participants in the neurology group reported statistically significant higher levels of overall satisfaction, particularly associated with the amount of time spent with clinical staff, consistency of care and information regarding the condition and its management.

4.4 Discussion

4.4.1 Overview

This section discusses the clinical and economic findings from the headache study. Strengths and limitations are also discussed, along with the potential implications for research and clinical practice in the management of chronic headache.

4.4.2 Rationale

Multiple studies concluded that advanced imaging, either CT or MRI, had the potential to change the clinical management of patients with chronic headache. The latter seems to be due to the reassurance effect associated with a brain scan with negative findings, particularly for brain cancer. Despite these findings and the regular utilisation of brain imaging in the NHS to reassure both GPs and patients alike, regulatory agencies (e.g. NICE) do not recommend the utilisation of advanced imaging in patients with chronic headache and no clinical red flags (i.e. for reassurance purposes). Probably due to this mismatch between official recommendations and real-world clinical practice,

little was known about the economic implications of providing GPs with direct access to neuroimaging for the management of chronic headache patients. The study's primary and secondary objectives hence considered the 6 and 12-month cost implications of providing GPs with direct access to brain MRI for the management of chronic headache patients compared to the conventional management to neurology services. Furthermore, secondary objectives considered cost-effectiveness and cost-utility analyses and other dimensions, from accessibility to care, patient satisfaction and diagnostic accuracy of both clinical pathways. These outcomes aimed therefore to evaluate the real-world implications of direct access to brain MRI compared to neurology services across a wide range of dimensions of care.

4.4.3 Key Findings

Primary outcome:

6-month cost analysis

These results are consistent with previous studies showing that direct access to advanced imaging resulted in a large decrease in referral rates to neurology services (Thomas et al. 2010; Howard et al. 2005).

The study's underlying hypothesis was that the early use of an advanced and accurate diagnostic tool (in this case MRI) would reassure both patients and GPs that no serious underlying cause (particularly brain tumour) is present. This would, in turn, reduce the headache burden and NHS resource use associated with the patient's subsequent management. Given the high prevalence of headache and the increased referral of patients with chronic headaches and other neurological conditions from primary care to hospital based care (Latinovic, Gulliford, and Ridsdale 2006), it is relevant to assess the implications of using different management strategies. To our knowledge, no previous prospective study in the UK assessed the economic implications of these two coexisting management strategies based on the GP referral decision.

The primary outcome considered the total headache-related NHS costs at six months post-recruitment. This time horizon was considered as most costs were anticipated to be realised within this timeline. However, due to increasing waiting times between events (e.g. time elapsed between the initial and follow-up neurology appointments), the 12 months' time horizon was found to be more appropriate to evaluate the cost implications of both clinical pathways. The study showed that the use of advanced imaging produced cost savings to the NHS compared to referral to neurology, with mean cost savings per participant of £333 and £518 at month 6 and 12 post-recruitment, respectively ($p<0.001$). This is equivalent to 57% and 64% cost savings per participant at, respectively, month 6 and 12 post-recruitment. These cost differences were multifactorial but primarily driven by: (i) the lower unit cost of a head MRI scan (£146) compared to the initial neurology appointment (£240); (ii) the lower number of outpatient appointments in the MRI group (mean 0.26 vs 2.52 appointments); (iii) the fact that 75/128 (59%) of participants in the neurology group ended up having a head MRI scan in the 12-month period of follow-up; and (iv) the increased

likelihood of patients in the neurology arm receiving management with Botox injections or nerve block procedures (mean 0.05 vs 0.30 events).

The results were similar to an uncontrolled clinical audit published by NICE (Wood 2018). Despite the methodology limitations of this study (e.g. only secondary costs seem to be considered, apparent 3-month time horizon, unclear as to whether the costs associated with management of incidental findings were included) and no formal report or paper could be retrieved, the use of direct access to MRI led to cost savings of 65% (£142k vs £410k) compared to management with referral to neurology services.

Direct access to head MRI seemed to reassure most participants as only 17 (18%) participants in the MRI group ended up being referred to a neurologist within the study timeline. This number confirmed the evidence from previous studies, such as Thomas et al. (2010). Furthermore over 66% of participants in the MRI group had no further hospital-based care, compared to only 5% in the neurology group. Similarly, at primary care level, participants in the MRI group had lower utilisation rates per participant when compared to the neurology group (mean GP appointments of 1.19 vs 1.82, p=0.006). All cost differences between groups remained statistically significant ($p<0.001$) when adjusted for differences in baseline characteristics or follow-up attrition rates.

A small proportion (7.6%) of participants in the MRI group had abnormal findings in the initial head MRI. However, only one participant had clinically significant lesions (two small aneurysms), which were incidental findings, and no brain tumour was diagnosed. This was consistent with clinical literature that estimated the prevalence of brain tumours among chronic headache sufferers to be less than 0.1% (Symvoulakis et al. 2007).

Secondary outcomes:

Cost-utility analyses

The *a priori* statistical analysis plan specified cost-utility analyses at 6 and 12 months post-recruitment. However, follow-up data from the EQ-5D-5L questionnaire had particularly poor participant compliance, with 12-month attrition rates of 75% and 85% of follow-up data missing in the neurology and MRI group, respectively. For this reason, it was decided only to conduct the 6-month cost-utility analysis. The 6-month cost-utility analysis showed that the use of direct access to head MRI a 7.2% and 4.1% of being cost-effective, taking, respectively, a £20,000 and £30,000 willingness-to-pay per QALY thresholds (typically considered by NICE). Despite saving money to the NHS, this very low probability was due to the lower utility values generated in the MRI group compared to the neurology group. However, as mentioned, these results are potentially affected by the high attrition rate in the MRI group, with a total of 73% (69/95) participants not responding to the EQ-5D-5L questionnaire at month 6. For this reason, and before this data are to be used by decision makers, further evidence concerning the validity of the utility values reported in both groups, particularly in the MRI group, should be considered.

Patient satisfaction

Patient satisfaction in both groups was compared based on three dimensions of analysis: during the referral period, the initial appointment and overall satisfaction. Twenty five percent of participants in the neurology group (vs 18% in the MRI group) reported dissatisfaction with the waiting time. This finding was not unexpected as the mean waiting time associated with the neurology appointment was almost three times of the one associated with the MRI scan (110 vs 39 days). Contrary to the referral period, participants in the neurology group had a trend towards satisfaction levels associated with the first appointment (neurology outpatient visit vs MRI scan) and a better experience compared to their expectations (60% vs 29% in the neurology and MRI group, respectively). Participants in the neurology group reported improved satisfaction levels at 3 months across different variables (time spent with clinician, consistency of care, information about the condition and its treatment). Almost three quarters of neurology participants reported being satisfied or very satisfied with their headache management compared to only 21% in the MRI group ($p<0.001$). The latter seems to be associated with the continuity of care provided by neurologists as opposed to the more fragmented care provided to participants in the MRI group. Both participants and GP referrers in the MRI group reported dissatisfaction associated with the waiting time elapsed between the MRI scan and the availability of results (mean time of 31 days). This may have contributed to an increased anxiety in some participants as opposed to the anticipated reassurance effect.

Headache burden

At baseline, participants in the neurology group reported lower quality of life and higher headache burden. HIT-6 scores improved over time but we were unable to assess whether there was a statistically significant difference in both groups due to high attrition rates, particularly in the MRI group.

As mentioned, out of the 95 participants recruited to the MRI group, 17 (18%) were subsequently referred to a neurologist. Interestingly, at baseline, these 17 patients reported higher headache burden compared to all 95 patients recruited to the MRI group (MIDAS score: 51.2 vs 44.8; MIDAS headache days: 55.8 vs 42.8). This finding seemed to suggest that these data might be useful to risk stratify patients and support GPs in their referral criteria. However, further research into this area is required.

Time off work

Time off work was also evaluated as a proxy of headache burden. Participants in the neurology group presented a trend ($p>0.05$) of higher number of days off work due to headache compared to participants in the MRI group. It is important to note though that a high proportion of missing data in both groups, particularly in the MRI group. Despite the summarised issues with the follow-up data, the current study did not seem to support the findings from previous studies that costs due to time off work represented up to 85% of the total costs associated with the management of chronic headache (McCrone et al. (2011).

4.4.4 Strengths, limitations and implications

Strengths

Several strengths associated with this study have already been discussed in the context of the scaphoid trial (subsection 3.4.5). This included a clear research question, a heterogeneous population reflective of clinical practice, two clear groups being compared and *a priori* statistical analysis plan. No potential allocation bias was present as referrers were not believed to have an interest – conscious or unconscious - in potentially biasing the study findings (most GPs were even unaware of the study). The estimate of NHS resource use data was primarily based on comprehensive and complete data retrieved from hospital-based databases that captured both the acute and elective elements of the pathway associated with the management of patients with chronic headache. These data were supplemented by both primary care utilisation data, collected from each participant's GP, and self-reported participant data. The objective was to ensure that all chronic headache related NHS events were costed regardless of the healthcare provider or setting. The prospective collection of healthcare utilisation and the evaluation of the impact of the interventions across different dimensions of analysis (efficiency, quality of care, access to care and patient satisfaction) were other key factors that contributed to the overall strength of the study.

Limitations

There were some limitations to this study. First, this was a single-centre study with participants recruited from one central Trust in London. A multi-centre study would be necessary to extend the generalisability of the results. Second, as with any observational study, no randomisation between groups was performed and there were significant differences in baseline of headache burden, health scores and utilisation of resources, reflecting potential selection bias. Although results were adjusted for potential confounders, this represented a study limitation. Third, specific inclusion and exclusion criteria were considered and, as such, the study sample might not be representative of all patients with chronic headache. Fourth, for the purpose of secondary outcomes, most data were self-reported and hence prone to recall bias. Lastly, there were high follow-up attrition rates potentially affecting secondary outcomes such as cost-utility analyses or patient satisfaction. A large proportion of participants failed to complete the headache diaries and follow-up headache-specific questionnaires even though these constituted part of their standard care. This was particularly the case in the MRI group as participants attended less often secondary care appointments. These large attrition rates could be minimised with closer cooperation between the clinical and the research teams in the proactive follow-up of participants.

Implications for Further Research

Baseline measures of headache burden, such as the HIT-6 or MIDAS, could potentially be used to determine which pathway may be suitable for individual patients but further research into the risk stratification of chronic headache patients is required (e.g. patients likely to be reassured by a negative brain MRI). Future studies should consider a closer engagement with participants in the MRI group, as these patients are less likely to attend secondary care services and comply with

study follow-up requirements. In addition, future research should further assess patient satisfaction associated with both clinical pathways.

Implications for Policy and Clinical Practice

With regards to clinical practice, the two clinical pathways are expected to face increasing demand and different transformation initiatives were identified based on the study findings. Chapter 7 discusses the different implementation initiatives developed to improve clinical practice for patients referred to GSTT with chronic headache.

The widespread use across the NHS of advanced imaging, particularly MRI, as an alternative to neurologist appointments for chronic headache sufferers holds the potential to, simultaneously, decrease overall NHS costs and release neurology resources to manage other clinical conditions (e.g. epilepsy, Alzheimer disease, dementias). However, it remains unclear as to what impact this management pathway will have on patient satisfaction or indeed outcomes such as QALYs.

4.5 Conclusion

This study found that the referral from primary care to direct access to head MRI compared to referral for a neurologist for patients with chronic headache was associated with lower NHS overall costs at 6 and 12 months post-recruitment. Despite waiting longer from referral to appointment, participants in the neurology group reported higher satisfaction levels associated with the care received compared to the MRI group and were more likely to benefit from changes to their therapeutic management.

The utilisation of direct access to MRI for a selected proportion of chronic headache patients, those more likely to be reassured, as a direct alternative to neurology services should be further evaluated and implemented within the NHS. This might ultimately contribute to the NHS financial sustainability whilst releasing neurology resources for other increasing prevalent neurological conditions. However, given the lack of follow-up utility data, further research should be consider to assess the cost-effectiveness of providing GPs with direct access to brain MRI for the management of chronic headache.

Chapter 5. Use of advanced imaging in the management of low to intermediate risk of suspected colon cancer

5.1 Introduction

5.1.1 Colorectal Cancer

Colorectal cancer (CRC) includes cancerous growths found in the colon, rectum and appendix. Most CRCs are the result of the malignant development of adenomatous polyps into invasive adenocarcinoma which occurs over many years (Riccioni et al. 2012, NICE 2011).

CRC is the third most common cancer in the UK (after breast and lung), with an annual incidence of approximately 40,000 new cases (NICE 2011a; Cancer Research UK 2015). The incidence of CRC is related to age, with almost three-quarters of cases occurring in people aged 65 or over and 83% in those over 60 years of age (NICE 2011a). For patients under 50 years of age, both sexes have similar rates, but later in life men tend to have an increased incidence of CRC (NICE 2011a). Over the past decade, overall CRC incidence rates remained relatively stable, with a small increasing trend (NICE 2011a).

CRC is, after lung cancer, the second most common cause of cancer death in the UK (NICE 2011a). Evidence gathered between 1997 and 2006 showed a decrease in excess of 30% in CRC mortality rates. This improvement seemed to be multifactorial but partly attributed to better detection and removal of colonic polyps, improved detection of colorectal cancer at an earlier stage, and the development of more effective primary and adjuvant treatments. In particular, early diagnosis of CRC is of vital importance as survival rates are stage-dependent at the time of diagnosis, with 5-year survival rates in men of 95% at stage 1 and less than 10% at Stage 4 (Cancer Research UK 2015). Despite this historical improvement, the UK 5-year survival rates for CRC remained significantly lower compared to other countries such as Sweden, Australia, Canada or Norway (Coleman et al. 2011).

The NHS Long Term Plan, published in the UK in January 2019, set out commitments to improve cancer outcomes and services in England over the next 10 years (NHS England 2019). One of its key ambitions was that, by 2028, the proportion of cancers diagnosed at stages 1 and 2 should rise from around 40% to 75% of cancer patients. Cancer alliances across the country have been established to implement the cancer strategy at a local level, with the introduction of faster, more standardised, diagnostic pathways. To further facilitate early diagnosis, the threshold for referral for patients with suspected colorectal cancer has been lowered (i.e. referral for colonic investigations in the presence of less severe symptoms such as change in bowel habits, regardless of its duration), incorporating updated NICE guidance (NG12) published in 2015 (NICE 2015).

5.1.2 The clinical challenge

Less than 10% of symptomatic patients referred to NHS outpatient clinics with suspicion of CRC are actually diagnosed with the disease (NICE 2011a). However, given the prevalence of the disease and the symptoms associated with it, the diagnosis of gastrointestinal diseases currently puts a considerable burden on secondary care. The lower threshold for referral for patients with suspected colorectal cancer has also led to an increase in demand for colonic investigations. The Department of Health has predicted an annual increase of 10% to 15% in the demand for colonic investigations, putting an additional burden on the already stretched optical colonoscopy capacity (Bowel Cancer UK 2019).

Optical colonoscopy (OC) is the diagnostic reference test for CRC but it is technically difficult and resource intensive and capacity constraints means that there are considerable waiting times for an OC test. In 2012, over 25% of NHS providers reported waiting times in excess of 4/6 weeks in 25%/5% of OC scans. In 2018, it was estimated that about half of NHS hospitals did not meet the 2-week target for urgent colonoscopy (Bowel Cancer UK 2019). These waiting times translate into delayed diagnoses and, potentially, poorer prognoses.

The utilisation of non-invasive CTC as a direct alternative to the gold standard OC has been introduced in routine clinical practice following clinical evidence of non-inferiority of Computed Tomography Colonography (CTC) in the diagnosis of medium to large polyps and CRC (NICE 2018a).

Clinical evidence: accuracy levels

The sensitivity of OC for the detection of CRC is estimated at 94.7% (95% CI: 90.4% - 97.2%) (Pickhardt et al. 2011) and for detection of large polyps (i.e. diameter > 10 mm) at 94% (Menardo 2004). CTC provides similar accuracy levels to OC for the detection of large and medium sized polyps and is particularly sensitive in the detection of symptomatic CRC (Halligan et al. 2005). Furthermore, systematic reviews showed that sensitivity and specificity of CTC improves with larger polyps (NICE 2011a). According to a meta-analysis performed by Halligan et al. (2005), CTC has a 93% sensitivity (95% CI: 73% - 98%) and 97% specificity (95% CI: 95% - 99%) for large polyps. In the diagnosis of CRC, CTC has a 95.9% sensitivity value (95% CI: 91.4% - 98.5%) (Halligan et al. 2005). This value is corroborated by Pickhardt et al. (2011), who estimated a sensitivity value of 96.1% (95% CI: 93.8% - 97.7%). It is relevant to highlight that older studies may underestimate the accuracy of CTC, as newer generation CT scanners are likely to improve the performance in the diagnosis of both large polyps and CRC.

Additionally, CTC has the potential to investigate other intra-abdominal organs that are not seen with OC. This may be an important benefit as it may reveal clinically relevant extracolonic conditions that might warrant treatment, but also a potential disbenefit, leading to further diagnostic and therapeutic procedures for conditions without clinical relevance.

In summary, CTC is highly sensitive, particularly for medium to large polyps and CRC. Given the relatively low prevalence of CRC in those referred with symptoms, CTC is increasingly being considered as a potential alternative to OC as a first line colonic investigation.

A drawback of CTC is that, in the event of positive findings, patients need to undergo a subsequent invasive technique for polyp removal and/or tissue biopsy. This means that these patients in effect undergo two instead of one diagnostic procedure. A multi-centre randomised trial estimated that 30% of patients had additional colonic investigations following CTC, compared to only 8% after OC (Atkin et al. 2013). However, it has been shown that less than 10% of symptomatic patients referred to NHS outpatient clinics due to suspected CRC are actually diagnosed with the disease (NICE 2011a). This means that CTC can potentially reduce the need for an invasive test in the majority of patients. A second disadvantage of CTC is associated with the use of ionising radiation and hence an increased probability of developing cancer in the medium to long-term. In contrast, CTC's non-invasive approach is an advantage as it minimises the risk of perforation (reported between 0.005%-0.03% with CTC compared to 0.06-0.19% with OC) (Berrington de Gonzalez, Kim, and Yee 2010). Furthermore, CTC's patient acceptability and overall satisfaction are likely to be improved by the use of a non-invasive imaging technology as opposed to the conventional OC (Halligan et al. 2007). Despite these disadvantages and advantages, both tests (CTC and OC) are deemed safe and are commonly used in routine clinical practice.

In a context of scarce resources, it is essential to take informed decisions on how to allocate the various human, technical and financial resources. This study aims to build on the existing evidence and to evaluate the clinical and cost implications of using CTC as a substitute test for patients referred from primary care with low to intermediate risk of CRC.

5.1.3 Two pathways at Guy's and St Thomas' Hospital

Patients with suspected CRC are commonly referred from primary care to hospital diagnostic services under a 'two week wait' referral. Following referral from primary care, the clinical pathway at GSTT includes the use of several imaging modalities, particularly OC and, to a lesser extent, CTC.

Optical Colonoscopy as the initial diagnostic test:

OC is the diagnostic reference test for CRC and hence the most frequently used initial diagnostic imaging tool at GSTT. A retrospective analysis of activity between January 2012 and November 2014 identified that GSTT performed 9,090 optical colonoscopies (260/60 procedures per month/week).

The subsequent diagnostic and treatment pathway is determined by the results from the OC (see Figure 57):

- **Positive** findings (i.e. abnormal findings on the initial optical colonoscopy):

Abnormal findings are subdivided into two categories: (i) medium to large polyps (over 6 mm), and (ii) CRC. For patients with medium to large polyps, polyps are biopsied and some of them removed. For patients with CRC, tissues are biopsied and appropriate cancer treatment options are then considered and agreed (e.g. chemotherapy, radiotherapy or surgery).

- **Negative** findings (i.e. normal findings on the initial optical colonoscopy):

No further diagnostic or treatment options are considered and the patient is discharged back to primary care (i.e. the referrer).

- **Inconclusive** findings (e.g. patient unable to comply with the procedure):

OC scan that does not produce conclusive results (e.g. suboptimal scan, incomplete bowel visualization). Usually, this is due to poor patient compliance prior to the procedure (i.e. patients did not follow bowel preparation processes) or during the procedure (e.g. unable to tolerate invasive scan). A multi-centre randomised trial reported a non-completion rate of 11% for OC (Atkin et al. 2013). For these patients, a repeated OC, flexible sigmoidoscopy or CTC scan are usually performed at GSTT.

Computed Tomography Colonography as the initial diagnostic test:

CTC utilisation at GSTT:

A retrospective analysis of activity between January 2012 and November 2014 identified that GSTT performed 1,642 CTC procedures (average of 47/11 procedures per month/week). These procedures were primarily performed on: (i) patients who had inconclusive findings in the initial OC and were then referred for CTC; (ii) frail patients who were not deemed fit for an invasive OC; (iii) as a direct alternative to OC.

Based on CTC's diagnostic accuracy and patient acceptability, as well as the increasing waiting times for OC, CTC has been increasingly used at GSTT as an alternative tool in the diagnostic pathway of symptomatic patients with suspected CRC (see Figure 57).

- **Positive** findings (i.e. abnormal findings on the initial CTC):

As with the OC group, abnormal findings are subdivided into two different categories: medium to large polyps (≥6mm) and CRC. For both findings, an invasive colonic test, either OC or flexible sigmoidoscopy (depending on the lesion's location), is subsequently performed in order to either excise (e.g. polyps) or biopsy lesions (e.g. CRC) identified on the initial CTC.

- **Extracolonic** findings (e.g. findings outside the colon):

Given the CTC's ability to visualise organs outside the colon, extracolonic findings include multiple clinical conditions, ranging from innocuous findings to potentially life-threating conditions (e.g.

abdominal aortic aneurysm). Based on the radiologist's report, patients are, if clinically relevant, subsequently referred to other clinical specialties for further assessment and/or treatment.

- **Negative** findings (i.e. no abnormal findings on the initial CTC):

No further diagnostic or treatment options are considered and the patient is discharged from secondary care.

- **Inconclusive** findings (e.g. patient unable to comply with the procedure):

Similar to OC, CTC might produce inconclusive findings (e.g. suboptimal scan, incomplete bowel visualisation). For these patients, a repeated CTC, a flexible sigmoidoscopy or an OC scan are usually performed at GSTT.

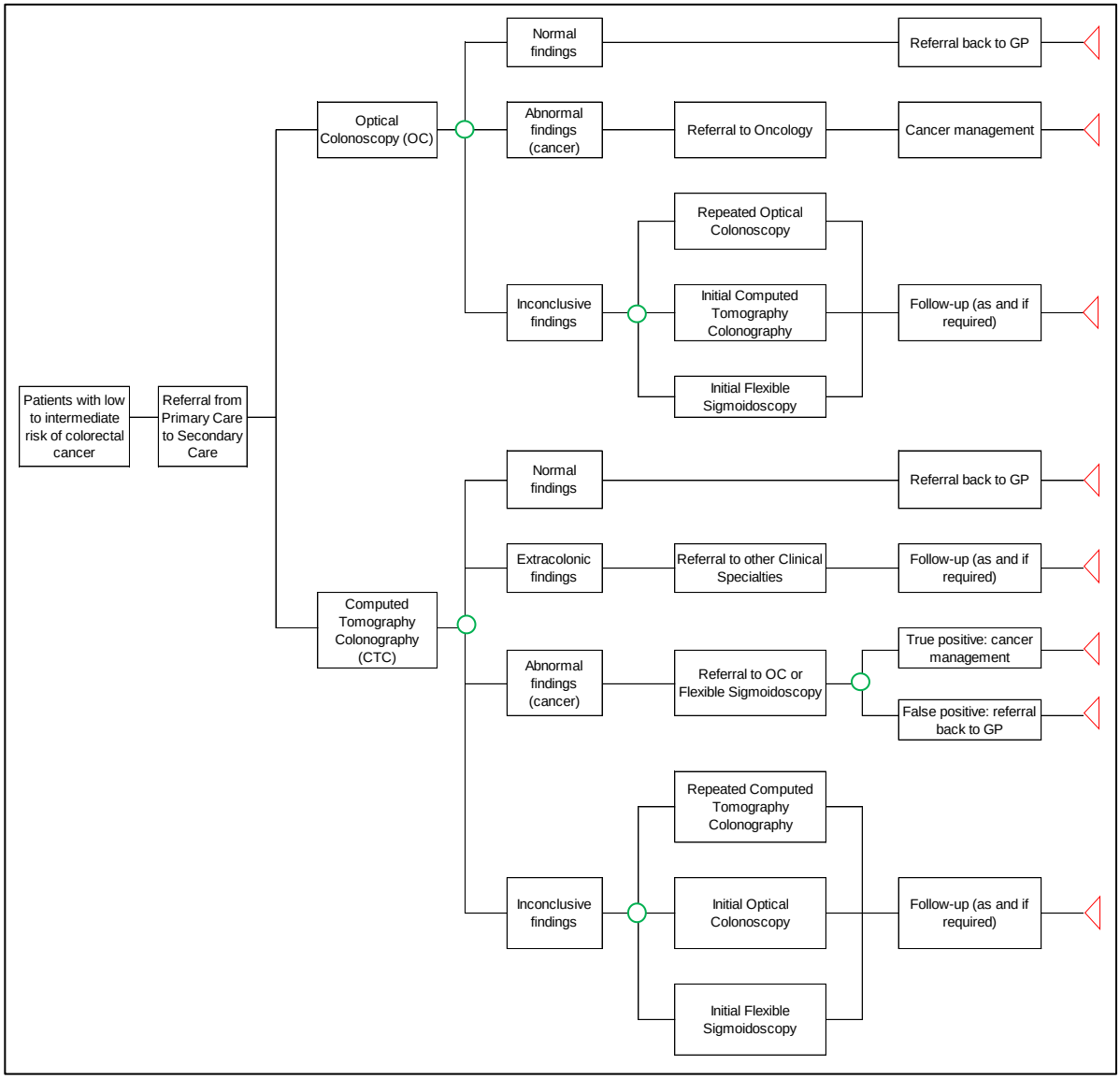


Figure 57. Clinical pathway associated with OC or CTC as the initial imaging modality for patients with low to intermediate risk of CRC.

5.1.4 Economic evidence

A review of literature was undertaken to evaluate the existing economic evidence around the economic impact of using CTC in the management of symptomatic patients with suspected CTC. Particular attention was paid to evidence retrieved from the NHS.

The vast majority of the cost analyses and economic evaluations performed assessed the role of CTC in asymptomatic patients, i.e. as a screening tool (Halligan et al. 2015). The economic evidence from screening studies was not considered in this chapter as it is unlikely to reflect the use of CTC in a symptomatic setting. The latter is due to multiple reasons. First, the prevalence of polyps and CRC is likely to be higher in symptomatic patients. Second, CRC diagnosed in the asymptomatic group are likely to be different (e.g. less aggressive) compared to those in the symptomatic group. Third, the sensitivity and specificity of CTC for the diagnosis of polyps and CRC are likely to be different in asymptomatic vs symptomatic patients. For these reasons, the evidence summarised below focused on the role of CTC in the diagnosis of symptomatic patients with suspected CRC.

Halligan et al. (2015) conducted a systematic review to evaluate economic evaluations on the use of CTC in asymptomatic and symptomatic patients. Out of sixteen papers, only one compared the use of CTC with OC in the diagnosis of symptomatic patients (Gomes et al. 2013). This study was a cost-utility analysis, which aimed to compare the utilisation of CTC and OC in symptomatic patients in the UK NHS using a Markov decision analytic model to estimate disease progression (Gomes et al. 2013). The authors concluded that the use of CTC was cost saving (mean costs per patient £467 vs £583) and “marginally” cost-effective (ICER of £23,000 per QALY), with a 60% probability of being cost-effective at the usual NICE willingness-to-pay thresholds values of £20,000-£30,000 per QALY) (Gomes et al. 2013). The remaining papers, while assessing the screening role of CTC, found that results were strongly influenced by the relative unit costs of CTC and OC and the inclusion or exclusion of costs associated with the management of incidental extracolonic findings (in the CTC group) (Halligan et al. 2015). A retrospective study of 225 CTC scans assessed the economic impact associated with the management of extracolonic findings and concluded that these approximately doubled the costs associated with the provision of CTC (Xiong et al. 2006).

As part of a large multicentre trial (SIGGAR trial, 1610 patients from 21 UK hospitals: OC group n=1,072, CTC group n=538), Halligan et al. (2015) performed a cost analysis around the utilisation of CTC in symptomatic patients compared to OC in the diagnosis of large polyps (≥10 mm) or CRC. This was the first randomised trial to directly compare the use of CTC as a replacement of OC as the initial diagnostic scan in the management of symptomatic patients. Based on a mean unit cost of £160 and £330 per CTC and OC scan, respectively, the authors estimated a total mean unit cost per participant of £674 for CTC and £739 for OC (Halligan et al. 2015). However, this £65 difference was not found to be statistical significant. Although cheaper than OC, patients undergoing CTC as the initial test were more likely to undergo a second investigation (30% vs 8.2%). In addition, 60% of patients undergoing

CTC had at least one extracolonic finding, 8.5% of whom requiring further investigations and around 2% had extracolonic malignancies detected. On the other hand, adverse events were more likely in patients undergoing OC, leading to a total of six admissions in the OC group compared to one in the CTC group. These three elements combined seemed to be the reason why there were no significant cost differences between groups were observed despite the unit cost of CTC being less than half than the unit cost of OC. As a corollary, the authors concluded that CTC was a safe alternative to OC in symptomatic patients, with higher patient satisfaction at a similar cost for the NHS (Halligan et al., 2015).

More recently, a retrospective study assessed whether constipation as a primary presenting complaint should be an indication for an OC or CTC scan (Ratnasingham et al. 2017). A total of 200 NHS patients were included in the study (100 in each group), with the authors reporting a 37% rate of incomplete OC (51% due to discomfort and 27% due to bowel preparation) and no incomplete CTC scans. Based on £434 and £518 unit cost for CTC and OC, respectively, the authors estimated a total cost per participant in the CTC group of £550 and £737 in the OC group. The authors concluded that CTC was cost-effective ($p<0.05$). However, there were several important limitations: the paper did not fully describe the NHS resource use considered in the cost analysis; it was unclear whether costs due to the management of extracolonic findings were included; and was uncertain as to what measure of effect was used in the cost-effectiveness analysis. Bearing in mind the study's retrospective design and the mentioned limitations, the authors concluded that CTC may be an acceptable alternative to CTC as a first-line investigation in patients presenting complaint of constipation (Ratnasingham et al. 2017).

In summary, three NHS studies evaluated the use of CTC among symptomatic patients. First, a cost-utility analysis, using a Markov model to estimate disease progression, concluded that the use of CTC was cost saving and marginally cost-effective (ICER of £23,000 per QALY), presenting a 60% probability of being cost-effective at typical willingness-to-pay values of £20,000-£30,000 per QALY) (Gomes et al. 2013). Second, a large multicentre randomised trial (SIGGAR trial) evaluated the cost implications of using CTC as a replacement of OC as the initial colonic investigation and found a statistically non-significant cost difference between groups (Atkin et al. 2013). Third, a retrospective study assessed whether constipation as a primary presenting complaint should be an indication for an OC or CTC scan (Ratnasingham et al. 2017). Despite the important methodological limitations associated with the cost and cost-effectiveness analyses, the authors estimated that CTC led to cost savings and was cost-effective ($p<0.05$). All three studies presented a common feature, the evaluation of CTC as a direct replacement of OC for patients with suspected CRC. Nevertheless, the participant's inclusion criteria was different among studies, with the first two studies considering a heterogeneous population of patients with suspected CRC whilst the third study focused on a subpopulation, patients presenting with constipation and a clinical suspicion of CRC. The decision concerning the inclusion criteria is essential as is likely to impact the incidence of CRC. As an illustration, patients with

constipation are less likely to have CRC compared to patients with multiple symptoms (e.g. constipation, rectal bleeding and anaemia). Given the non-invasive nature of CTC, its added value resides mainly in its ability to rule-out CRC or medium to large polyps without the need for an invasive technique (e.g. OC or flexible sigmoidoscopy). In fact, patients undergoing CTC that are found to have CRC or large polyps routinely undergo an invasive procedure for biopsies and/or therapeutic excision of polyps. Hence, the more prevalent polyps and CRC are, the less likely the intervention with CTC will lead to cost savings. Based on this rationale, we have chosen to evaluate the use of CTC as the initial colonic investigation deliberately among patients with low to intermediate risk of CRC. The study's inclusion criteria included adult patients over 40 years old presenting with constipation or change in bowel habits and excluded patients with anaemia, rectal bleeding, with diarrhoea only for more than 6 weeks, among others.

5.1.5 Patient and NHS benefits and disbenefits

It was anticipated that the increased utilisation of CTC could deliver several benefits including:

- Improved access to care by reducing waiting times for colon imaging;
- Faster diagnosis and, if needed, subsequent treatment that could lead to better clinical outcomes;
- Diagnosis of potential extracolonic findings that are clinically relevant and would not otherwise be picked up by using OC;
- Improved patient satisfaction as CTC is a non-invasive scan and does not involve the same level of discomfort or potential risks (e.g. bleeding) associated with optical colonoscopy;
- Increased patient compliance with optical colonoscopy for those with positive CTC findings;
- Freeing up OC capacity to perform procedures in high-risk patients;
- Potential reduction of overall cost per symptomatic patient.

The increased utilisation of CTC was thought to be associated with the following disbenefits:

- An increase in the number of patients who may need to undergo two diagnostic tests (CTC and subsequent optical colonoscopy) instead of one, as will happen for those patients where there is a major positive finding (either CRC or medium to large polyps) on the initial CTC.
- Increased exposure to ionising radiation and, therefore, an increased risk of cancer induction, particularly in the medium to long-term;
- Diagnosis of potential extracolonic findings that might not be clinically relevant and would not otherwise be picked up by using OC.

5.2 Methods

5.2.1 Aims, Objectives and Hypotheses

Aim of the study

The aim of this study was to investigate whether the diagnosis of patients with suspected CRC using CTC as the first diagnostic imaging tool is cost saving when compared to OC.

Given the similar accuracy levels and waiting times associated with both imaging modalities, it was not deemed likely that the initial use of OC or CTC would be associated with a CRC stage shift. For this reason, the primary objective aimed to estimate the cost difference between the OC and CTC groups.

Primary Objective:

To evaluate whether the use of CTC as the first imaging modality for a selected subset of patients with suspected colon cancer is cost saving at 6 months post-recruitment compared to the utilisation of OC as the first imaging modality. The primary analysis included all costs associated with the management of colon-related events as well as the management of potential extracolonic findings (incidental or non-incidental findings) in participants of the CTC group. Costs associated with the treatment of cancer were not considered as part of the baseline analysis so not to skew the analysis due to potential imbalances in cancer detection rates between both groups. However, cancer treatment costs were considered in a secondary cost analysis.

Secondary Objectives:

Seven secondary objectives, together with a short rationale, are presented below.

Objective 1: To evaluate whether the use of CTC as the first imaging modality was cost-effective at 6 months compared to the utilisation of OC as the first imaging modality.

Cost and utility data at 3 and 6 months of follow-up were used to assess the intervention's cost-effectiveness. The final 6-month follow-up following the initial imaging test (either OC or CTC) was considered appropriate as all relevant costs and outcomes associated with both clinical pathways should be realised within this timeline. Outcomes were expressed in QALYs based tariff weights derived from the EQ-5D-5L questionnaire. It was assumed a linear relationship between two health states (i.e. the QALY between 0-6 months is the average between these two points). In order to adjust for imbalance in mean baseline utilities, a multiple regression analysis was used. The cost-utility analysis, which is the preferred method of economic evaluation of NHS interventions, was performed according to NICE recommendations (NICE 2012a).

Objective 2: To evaluate the cost per correctly diagnosed CRC using CTC as the initial investigation compared with OC.

The second objective was to evaluate the cost per correctly diagnosed medium to large polyps (≥6mm) and colorectal cancer using CTC as the initial colonic investigation.

For patients with positive findings on CTC (either medium to large polyps and/or CRC), the subsequent invasive scan (OC or flexible sigmoidoscopy) was used as the comparator to assess whether the initial diagnosis with CTC was correct.

Patients with negative findings on CTC were followed-up for a period of 12 months. In order to identify any cancers missed, all patients with negative findings for CRC were cross referenced against the NHS Central Register (NHSCR) and details of new cancer diagnoses and deaths obtained from NHSCR. Patients were also matched with national data from the Hospital Episode Statistics (HES) database to reduce the time lag between cancer diagnosis and the time of notification. A CRC diagnosis was defined as missed if was identified through one of the databases and no diagnosis of CRC was present on the initial CTC test.

Objective 3: To estimate the incidence of extracolonic findings in the CTC group.

This objective aimed to assess the number of extracolonic findings and the proportion of those who warranted further diagnostic or therapeutic management.

It was anticipated that significant extracolonic findings would be diagnosed based on CTC. Some extracolonic findings (e.g. abdominal aneurism and extracolonic cancers) could be life-threatening and therefore likely to require treatment within the study's six month follow-up period, whilst others may not warrant any further diagnostic or therapeutic management.

The presence or absence of extracolonic findings were retrieved from the original CTC report. In the event of presence of extracolonic findings, potential subsequent diagnostic or treatment care was assessed for a period of six months based on primary and secondary care databases.

Objective 4: To compare the likelihood ratio of CTC against colonoscopy in patients with initial positive CTC findings for medium to large polyps (≥6 mm) and CRC.

The likelihood ratio of CTC against colonoscopy in patients with initial positive CTC findings were assessed using the invasive scan (OC or flexible sigmoidoscopy) as the gold standard.

As per standard care, all patients with positive findings on the initial CTC underwent an invasive test, either an OC or optical flexibility. Findings from both tests were compared using the invasive test as the reference. The likelihood ratio for a positive CTC was estimated based on the equation below.

$$\text{Likelihood ratio (LHR) for a positive CTC} = \frac{\text{True Positive}}{\text{False Positive}}$$

Objective 5: To measure the time taken to reach a definitive diagnosis and first major treatment decision based on the CTC findings in comparison to the current pathway.

This objective evaluated access to care for both groups, measuring the time elapsed between the day of referral to either OC or CTC and: (i) the day the test was actually performed; (ii) the day a definitive diagnosis was reached (OC or flexible sigmoidoscopy were considered the gold standard for patients presenting positive findings in the initial CTC); and (iii) the day a treatment decision was reached. For this objective, three time points were necessary: date of referral from primary care for a diagnostic test (retrieved from referral form); date of diagnostic test (coincident with recruitment date); and date at which the OC or CTC report is available to the referrer (retrieved from primary and secondary care databases).

Objective 6: To compare the 'on the day cancellation' rates and incomplete (or suboptimal) bowel investigations associated with OC or CTC as the initial imaging test.

This objective assessed the operational efficiency associated with the provision of both tests by comparing the 'on the day cancellation' rates for either OC or CTC. Given that the bowel preparation associated with CTC is thought to be easily tolerated, it was anticipated that the on the day cancellation and incomplete bowel investigation rates would be lower in the CTC group. This, in turn, would reduce the need for subsequent bowel investigations.

The information concerning incomplete bowel investigations was obtained from the OC and CTC reports.

Objective 7: To compare patient satisfaction levels associated with the use of OC as the initial imaging modality in patients with suspected CRC compared to CTC.

This objective aimed to assess patient satisfaction in both groups using a participant questionnaire the morning after the bowel test. The questionnaire were mainly based on a 5-point Likert scale, used to assess any potential differences between the two groups. The design of the post-test patient satisfaction questionnaire was based on the RCT conducted by Atkin et al. (2013). The underlying hypothesis was that the use of CTC would promote an improvement in the participant's overall satisfaction, both at short and medium-term, compared to OC as it would decrease: (i) patient anxiety associated with an invasive procedure; and (ii) the time taken to obtain a definitive diagnosis.

Hypotheses

Null Hypothesis

There is no statistical difference between the six months cost per patient for those referred with suspected colorectal cancer to either: (i) an OC scan; or (ii) a CTC scan.

Alternative Hypothesis

There is a statistical difference between the six months cost per patient for those referred with suspected colorectal cancer to either: (i) an OC scan; or (ii) a CTC scan.

5.2.2 Study design

The study was an independent, single-centre site, observational study comparing the utilisation of CTC with OC as the first diagnostic tool in patients with suspected CRC. The study compared two existing clinical pathways used in the management of patients with low to intermediate risk of colorectal cancer following referral from primary care that differed in the initial diagnostic test with either CTC or OC (Figure 57).

As highlighted in Figure 58, the length of time each patient participated in the study was six months. Given the expected short-term impact of the intervention (diagnostic scan), it was considered that the majority of costs and outcomes were captured within the proposed six month follow-up (as highlighted in Figure 58 and Table 62). In addition, all patients were matched against NHSCR or the HES databases to determine whether a diagnosis of cancer (CRC or other cancer) had been made in the 12 months following the initial CTC or OC test.

Intervention groups and allocation

Participants were allocated to two groups: (i) the OC group; or (ii) the CTC group. The allocation was decided during a telephone clinic with a specialist nurse following referral from primary care, taking into account the patient's clinical history (e.g. clinical symptoms, overall health condition and previous colonic investigations) as well as her/his test preference. Given the potential selection bias, the primary outcome was adjusted taking into consideration potential imbalances in baseline characteristics (detailed in section 5.3.4).

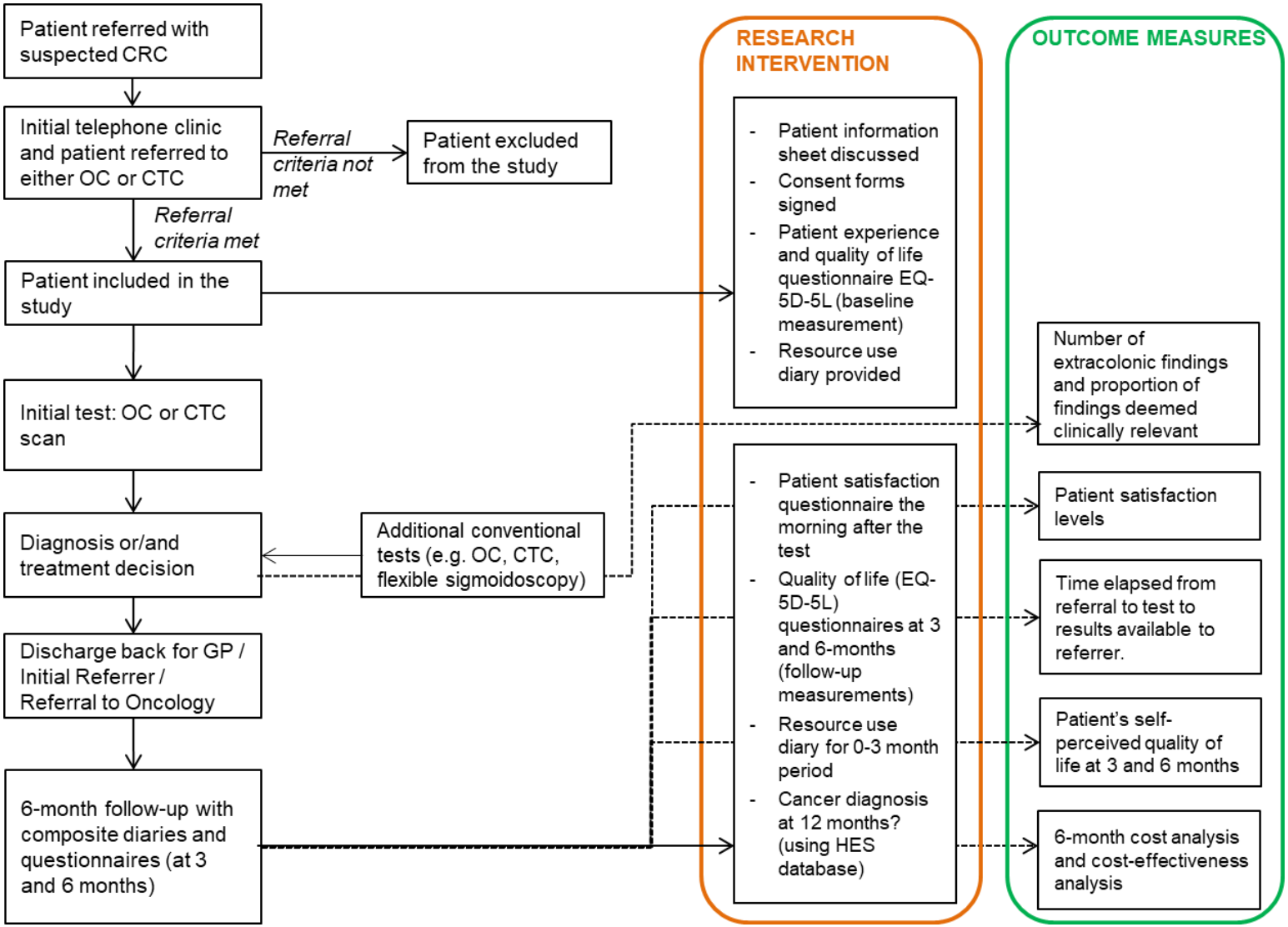


Figure 58. Colorectal cancer study structure.

Table 62. Study flowchart for both groups of patients.

Activity	Timing of activity							
	Prior to registration for study	Immediately after registration into the study	During OC or CTC scan	After OC or CTC scan				
				Next Day	3 months	6 months	12 months	Responsible
Give patient information sheet, explain study and obtain signed informed consent	X							GSTT
Register patient into the study	X							GSTT
Register patient demographics and clinical history		X						GSTT
Give patient registration pack (patient information sheet, copy of informed consent)		X						GSTT
OC or CTC scan			X					GSTT/Patient
Baseline questionnaires (EQ-5D-5L)		X						GSTT/Patient
Test's acceptability questionnaire				X				GSTT/Patient
3 and 6-months Follow-up questionnaires (EQ-5D-5L)					X	X		GSTT/Patient
3-months Follow-up Questionnaire (Patient Experience)					X			GSTT/Patient
Resource use diary					X			GSTT/Patient
Access NHSCR or the HES databases							X	GSTT

Follow-up Period

All participants were followed up for 12 months following the recruitment in the trial. Data were collected at baseline, 3 and 6 months post-recruitment. In addition, at 12 months post-recruitment, all participants were matched against a cancer registry and hospital databases to assess whether a diagnosis of cancer (colorectal or otherwise) had been given.

5.2.3 Ethical Approval, Trial Registration and Funding

The Health Research Authority and Research Ethics Committee (East Midlands – Leicester Central Research Ethics Committee) approved the study research on 04th April 2016. The REC reference was 16/EM/0143 and the study was registered on clinicaltrials.gov (Clinical Trial Registration: NCT02820389). The first participant was recruited on 14th June 2016.

The study was fully funded by a grant secured from Guy's and St Thomas' Charity.

5.2.4 Selection, withdrawal of participants and sample size

Study Setting

Patients were recruited at GSTT – either in the Colorectal Unit or the Radiology Department – as per their standard care. There were no changes to routine clinical practice in both groups as the study focused on evaluating two established clinical pathways in the management of patients with suspected CRC.

Inclusion Criteria

Patients considered to be suitable for the study included every adult patient over 40 years old, presenting with constipation or change in bowel habits with a differential diagnosis of suspected colorectal cancer.

Exclusion Criteria

Patients were considered to be ineligible for the study if at least one of the exclusion criteria was present:

- Patients presenting with anaemia;
- Patients presenting with diarrhoea only for more than 6 weeks;
- Patients presenting with rectal bleeding;
- Patients who had undergone a previous whole-colon examination in the past 6 months;
- Patients who had been referred for a whole-colon examination to follow-up a known colorectal cancer;
- Patients lacking capacity to give consent or participate in the study;
- Patients already taking part in any clinical trial of an investigational medicinal product;
- Patients who were not fluent in English.

Sample Size

The sample size estimate was calculated based on the primary endpoint, total colon-related six months healthcare costs. A total of 110 participants were recruited in the OC group and 70 participants in the CTC group to achieve a detection a cost difference of £200 assuming standard deviations of £400 and £300, respectively, with 90% power at the 5% two-sided significance level. A 20% increase in sample size was used to account for unknown cost distributions and attrition rate.

Losses to Follow-Up

If a participant moved from the GSTT catchment area, every effort was made to ensure the participant was still followed up. If a participant was lost to follow-up, individual patients' GP were contacted to obtain information. If a participant requested to be withdrawn from the study, the participant's data were excluded from the analysis.

End of Study

For regulatory purposes, the end of the study occurred 12 months after recruitment of the final patient at which point the 'declaration of end of trial' form was submitted to ethical committees. The last participant (180th participant) was recruited on 05th December 2018.

5.2.5 Data Collection and Outcomes at baseline and follow-up

Data were collected by a research team member at baseline and then quarterly up to 12 months following recruitment. Data collection was completed in December 2019.

Data at baseline were collected before the initial appointment (either CTC or OC scan) at GSTT. Follow-up data were collected as per the participants' preference, either via phone, email or post. Participants' preferences were established at baseline and recorded in the Case Report Form.

Participant Demographics

A variety of information was captured at baseline, including:

1. Age;
2. Gender (male/female);
3. Number and type of bowel related complaints;
4. Number and type of comorbidities / active health problems. This information was obtained from participants' self-reported data, data from primary care (classified under 'active problems' in the patient's electronic record) and secondary care clinic letters.
5. Number of months with change in bowel habits (numerical variable);
6. Healthcare utilisation in the 6 months prior to study recruitment;
7. Quality of life questionnaire (generic questionnaire: EQ-5D-5L).

Perspective of Analysis

The study took a National Health Service perspective of analysis over a six months' time horizon following recruitment. Only costs of all colon-related NHS events were considered. This approach is consistent with the methodology recommended by NICE for the evaluation of interventions with potential impact on health outcomes (EUnetHTA 2015).

The estimate of the total costs from a NHS perspective was based on the multiplication of: (a) any colon-related healthcare events; by (b) the unit cost of such events.

Resource Use Measurement

Resource use data included contacts with NHS healthcare providers associated with the management of the suspected colorectal cancer. In addition, given the CTC's ability to visualise extracolonic findings, any healthcare event due to an incidental finding was also included in the cost analyses. More details on the methodology used to measure resource use data were already described in the scaphoid and headache chapters (refer to sections 3.2.6 and 4.2.6).

Valuation of Unit Costs

Table 63 lists the unit costs associated with the primary outcome, including the reference and the rationale behind any assumption.

For the purposes of the primary outcome, the valuation of unit costs was, whenever possible, based on National Reference Costs 2016-17 (NHS Improvement 2017). All secondary care contacts were costed using this strategy.

For primary care events, an average cost per appointment (e.g. GP face-to-face appointment, GP phone appointment) was derived from the Unit Costs of Health and Social Care 2016 and then inflated to 2017 using the hospital & community health services (HCHS) index (Curtis and Burns 2017).

No medication costs were considered as differences between the two groups were not anticipated. Additionally, medication costs were mainly associated with over-the-counter painkillers (e.g. paracetamol, ibuprofen, buscopan) and such out-of-pocket expenditures are outside the NHS perspective of analysis.

If a scan, either a CTC or OC scan was not performed on the day (e.g. due to poor bowel preparation in both tests and unexpected claustrophobia in the CTC scan) a 50% of the unit cost was applied.

Table 63. Unit costs for all primary and secondary care events considered in the colon study.

Category	Unit Type	Unit cost (£)	Reference
Primary care			
GP appointment (face-to-face)	Per appointment	£36.50	Unit Costs of Health and Social Care 2016 and inflated to 2017 using the HCHS index (Curtis and Burns 2017).
GP phone appointment	Per appointment	£14.80	
Secondary care			
ED episode due to bowel presenting complaint	Per episode	£222	Reference Costs 2017 (NHS Improvement 2017).
Initial Telephone Triage clinic	Per appointment	£100	
Initial Outpatient appointment	Per appointment	£240	
Follow-up Outpatient appointment	Per appointment	£92	
CTC test	Per scan	£250	
Optical Colonoscopy	Per scan	£515 ; £654; £911	Diagnostic Colonoscopy 19 years and over (FZ51Z HRG code); Diagnostic colonoscopy with biopsy ≥ 19 years (FZ52Z); Therapeutic Colonoscopy 19 years and over (FZ53Z HRG code); Combined Upper and Lower GI Tract Diagnostic Endoscopic Procedures with Biopsy, 19 years and over (FZ64A HRG code)
Flexible sigmoidoscopy	Per scan	£398; £479	Reference Costs 2017, Radiology section (NHS Improvement 2017).
MRI scan (1 zone)	Per scan	£146	
CT scan (2 zones)	Per scan	£155	
Ultrasound	Per scan	£55	

5.2.6 Statistical Analyses

Analysis Population

All analyses were based on the principle of intention-to-treat (ITT).

Data Cleaning and Data Validation

All baseline and follow-up data cleaning was performed prior to any data analysis.

Baseline data were captured via a paper-based Case Report Form (CRF) during recruitment and then entered into a web-based CRF. During this process, the PhD student screened the data looking for inconsistencies. In the event of any potential data errors in the original hand-written data packs, participants or members of the research team were asked for clarification (e.g. date of birth and age did not match) and amendments were made to the original data set (data editing).

The NHS resource use measurement considered in the primary outcome was derived from the merge of medical records databases and data self-reported by participants. This comprehensive data collection methods (already detailed in subsection 3.2.6 and 4.2.6 for the scaphoid and headache studies) considered the validation of data using multiple datasets.

Missing Data

Participants were not excluded from the analysis due to any missing data, particularly as data related to the primary outcome was expected to be complete given the comprehensive data collection methodology. Only data from participants who withdrew their informed consent were not included in the analyses. However, where data for the estimate of the total healthcare costs were missing (e.g. missing primary care data), mean values from the respective group were imputed.

Baseline comparability of groups

Continuous data were summarised by: frequency, mean, standard deviation, minimum, first and third quartile, median and maximum. Tabulations of frequencies for categorical data were presented, as well as the percentage (%) relative to number of non-missing values within the respective intervention group, unless otherwise specified.

Given the non-randomised study design, significance testing was performed on the baseline variables between intervention groups listed under subsection 5.2.5 Data Collection and Outcomes at baseline and follow-up. Chi-square test was used to assess categorical variables. Quantitative variables were tested for normality using the Shapiro-Wilk test and, depending on this result, independent t-test or Mann-Whitney U test analyses were performed. The Levene's test was used to assess the homogeneity of variance [consistent with Peacock, Kerry, and Balise (2017)]. A p-value of $p<0.05$ was deemed as statistically significant.

Primary and secondary objectives

This study was observational but all analyses were based on the of 'intention-to-treat' so that all participants recruited were included in the analysis as per the group they were recruited to,

regardless of whether they actually received the intended intervention or not. Given the study's time horizon of 6 months, no discounting of costs or effects was considered.

Baseline sociodemographic and clinical characteristics were compared in the two intervention groups, i.e. gender, age, number of bowel-related complaints, number of active health problems/co-morbidities, number of months with change in bowel habits and self-reported quality of life (EQ-5D-5L). T tests were used for continuous data assuming a normal distribution, the Mann Whitney U test was used for non-normal quantitative data and categorical data were compared using chi-squared tests.

Given the skewness associated with the cost distribution, all cost differences between groups were assessed using GLM with an identity-link and gamma distribution. An identity link function instead of a log link was used to give estimates as means to avoid potential analytical biases (Polgreen and Brooks 2012; Barber and Thompson 2000). An unadjusted GLM cost analysis with the study group (OC group vs CTC group) as the only covariate was performed as the first step. Given the study's observational design and potential selection biases during the initial telephone clinic appointment, the groups being compared could have potentially been different due to the lack of randomisation with subsequent impact on healthcare costs (Moran et al. 2007). For this reason, an adjusted analysis was performed including all baseline variables with $p<0.10$. For all GLM analyses, group difference estimates and associated confidence intervals were reported, together with p-values.

The incremental analysis of effectiveness considered QALYs as the measure of effect. EQ-5D-5L questionnaires (EuroQol Research Foundation 2019) at four points in time (baseline and 1, 3 and 6 months post-recruitment) were used to generate quality of life scores (QALYs) based on a UK study (Devlin et al. 2018) and using area under the curve methods assuming linear movement between adjacent points (Drummond et al. 2004). A multiple regression analysis was used to address the potential imbalance between utilities at baseline likely to be correlated with the QALYs over the follow-up period (Manca, Hawkins, and Sculpher 2005). If utility data were missing, multiple imputation methods were used to assess the assumption that the data were missing at random. Missing data were imputed using 'multiple imputation using chained equations' (MICE), with the number of multiply imputed data-sets to be equal to the fraction of incomplete service-use information (White, Royston, and Wood 2011). 1000-replicate bootstrap analyses showing difference in costs and outcomes were presented on cost-effectiveness planes. All analyses were conducted using Stata version 15 for Windows (StataCorp LLC, Lakeway Drive,Texas).

5.3 Results

5.3.1 Data Validation and Completeness

Only participants that withdrew the informed consent and hence were considered lost to follow-up, were not included in the data analysis.

Data at baseline were complete for all participants, except for the number of comorbidities / active health problems (missing 5 and 2 participants in the OC and CTC group, respectively) and the number of months with change in bowel habits (missing 7 and 2 participants in the OC and CTC group, respectively).

Follow-up data were captured via the combination of data from primary and secondary care databases and self-reported data from participants. Data from secondary care databases was 100% complete (n=105 for the OC group, n=68 for the CTC group). Data from primary care databases was 98% complete (n=102 for the CTC group, n=67 in the CTC group). In the absence of data from both the primary care databases, any resource use outside GSTT was missing. Missing resource use values were imputed using the mean values from the respective group.

Data from the EQ-5D-5L questionnaire at baseline and 6 months post-recruitment were, respectively, 100% complete (n=105 for the OC group, n=68 for the CTC group) and 59% complete (n=58 for the OC group, n=44 for the CTC group).

5.3.2 Participant Flow

Participant flow associated with the colorectal cancer study is illustrated in Figure 59. A total of 180 participants were recruited, 110 to the OC group and 70 to the CTC group. During the follow-up duration, 4.5% (n=5) and 2.9% (n=2) of participants withdrew the informed consent in the OC and CTC group, respectively, and therefore were considered lost to follow-up. All participants that did not withdraw informed consent (n=173) were included in the analysis, 105 and 68 participants in the OC and CTC group, respectively.

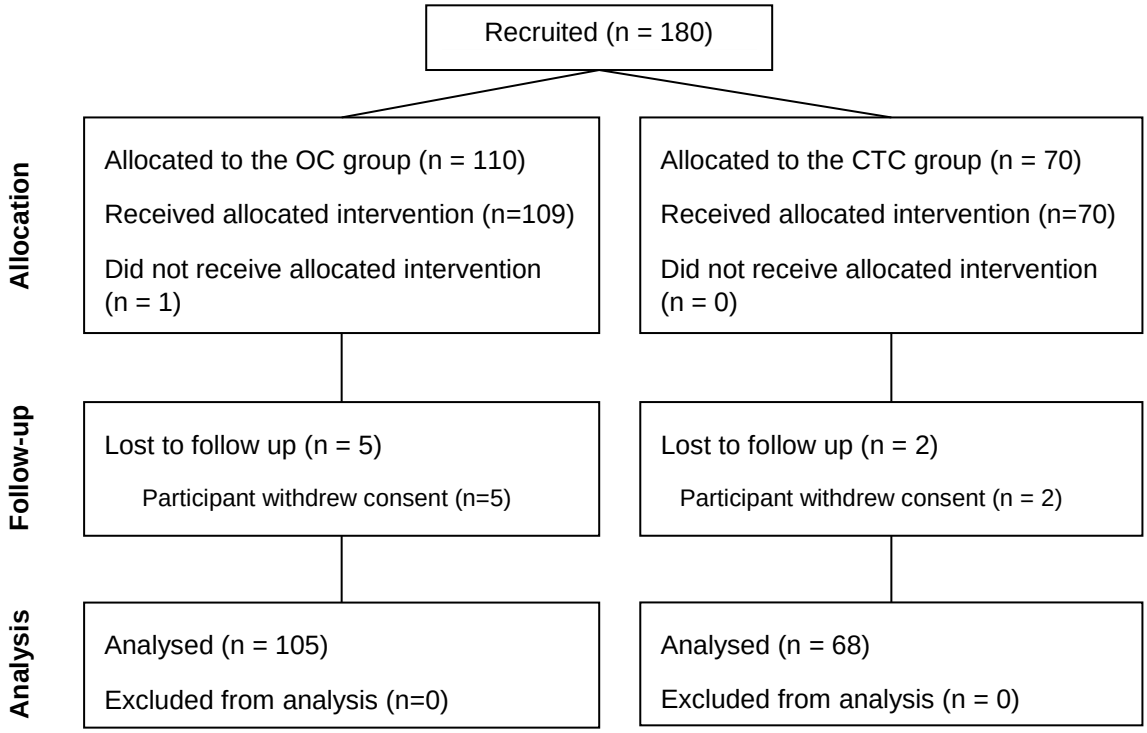


Figure 59. Participant flow chart for the colon study.

5.3.3 Participant Characteristics – Baseline

Table 64 describes the baseline sociodemographic and baseline outcome variables organised by study group. All participants, apart from those who withdrew consent, were included in the baseline analysis (n=173), distributed in the OC group (n=105) and the CTC group (n=68).

As summarised in the Statistical Analysis Plan section (section 5.2.6), significance testing between intervention groups was performed given the non-randomised study design.

Table 64. Baseline characteristics of the population by study group.

Variable	OC group (n=105)	CTC group (n=68)	p-value
Age, years: mean (SD) *	61.3 (10.4)	69.2 (11.4)	<0.001
Gender Female, n (%)	54 (51)	36 (53)	0.846
Number of comorbidities / active health problems: mean (SD) *	2.9 (2.1)	4.0 (2.2)	0.002
Number of months with change in bowel habits: mean (SD)	3.7 (3.9)	4.1 (3.7)	0.465
Self-reported health-related quality of life EQ-5D-5L: mean utility at baseline (SD) *	0.832 (0.197)	0.793 (0.173)	0.023
Total NHS costs in the 6 months prior to recruitment: mean (SD)	£187 (£65)	£206 (£155)	0.242

SD= Standard deviation; EQ-5D-5L - 5 level EQ-5D; * denotes a statistically significant difference between groups.

A similar proportion of females were recruited to both groups but participants in the CTC group were older, with a mean age of 69.2 years old compared to 61.3 in the OC group (p<0.001). In terms of clinical characteristics, participants in the CTC group reported greater disease burden levels with: (i) higher mean number of comorbidities / active health problems (4.0 vs 2.9 in OC group, p=0.002); and (ii) lower mean utility values at baseline (0.793 in the CTC group vs 0.832 in the OC group, p=0.023). With regards to NHS resource use in the 6 months prior to recruitment to the study, no statistically significant difference between the two groups was found.

Clinical findings:

Out of the 105 participants in the OC group, 5 cancers (4.8%) were diagnosed during the 12-month follow-up period. Three of these cancers were CRCs and diagnosed as part of the initial OC test

(sigmoid colon T3 N2; descending colon T4a N1b; anal canal T3 N0). Out of the 68 participants in the CTC group, a total of 6 cancers (8.8%) were diagnosed, with four being CRCs (2 polyp adenocarcinomas extracted; rectal cancer T3 N1 and colon cancer T4b N1c M1) diagnosed during the initial CTC. With regards to medium to large polyps, a higher proportion of participants in the OC group had one or more polyps diagnosed (40% compared to 13% in the CTC group).

5.3.4 Primary Objective

The primary objective was to estimate the 6-month colon-related costs associated with both groups. For this purpose, all participants were followed-up for a period of six months to capture all relevant NHS resources used in the management of suspected CRC patients. This included both primary and secondary care resources. Table 65 summarises the mean number of NHS events per participant for both groups. With regards to primary care appointments, participants in both groups presented similar utilisation rates of both GP face-to-face and phone appointments. If secondary care hospital appointments were considered, participants in both groups had similar utilisation rates, except for the diagnostic tests. Out of the 105 participants in the OC group, 12 (11%) had subsequent investigations, with 5 being repeated OC, 1 flexible sigmoidoscopy and 6 CTCs. In the CTC group, 22 (32%) participants underwent additional invasive testing, equally split between 11 OC and 11 flexible sigmoidoscopy. Participants in the OC group had a higher utilisation of OC scans ($p<0.001$) whilst patients in the CTC group underwent more CTC and flexible sigmoidoscopy scans ($p<0.001$).

Table 65. Breakdown of number of NHS appointments per type of activity organised per group.

	OC group (n=105)		**CTC group (n=68)**		**p-value**
Type of NHS appointment	**Total of episodes**	**Mean**	**Total of episodes**	**Mean**	
Primary care services					
GP face-to-face appointment	59	0.56	41	0.60	0.762
GP phone appointment	12	0.11	17	0.25	0.155
Hospital based services					
ED visit	3	0.03	0	0	0.083
Outpatient appointments	22	0.21	14	0.21	0.001
CTC scan	6	0.06	68	1.00	<0.001
OC scan	111	1.06	11	0.16	<0.001
Flexible sigmoidoscopy	1	0.01	11	0.16	<0.001

The mean cost of management per participant [mean (SD)] was lower in the CTC group compared to the OC group [£645 (£607) vs £991 (£316)], leading to an unadjusted mean cost difference between groups of -£345 per participant (p<0.001) (Table 66). Hence, at 6 months, there was a statistically significant difference between the two groups.

Table 66. Unadjusted and adjusted GLM analyses of costs 6 months post-recruitment.

Mean total cost (SD)	OC group (n=105)	CTC group (n=68)	Unadjusted cost difference (CTC group – OC group; 95% CI, p-value)	Adjusted cost difference (CTC group – OC group; 95% CI, p-value)
6 months	£991 (£316)	£645 (£607)	- £345 (-£501 to -£190) p<0.001	-£370 (-554 to -£185) p<0.001

The cost distribution is positively skewed (mean >> median), as it was affected by a small proportion of patients that have significantly higher costs (maximum cost of £2,166 and £1,864 for the OC and CTC group, respectively). The cost difference between groups was mainly driven by the higher proportion of participants in the CTC group in the £0 to £500 range compared to the OC group [41 (60%) vs 0 participants] (Figure 60). At the opposite end, a marginally higher proportion of participants in the CTC group [5 participants (7.4%) vs 6 (5.7%)] had very large costs (>£1,500). The latter was mainly due to participants that required more than one diagnostic scan and costs associated with the management of extracolonic incidental findings. As an example, one participant was found to have a mass in the omentum, a situation that led to its surgical excision with a laparotomy.

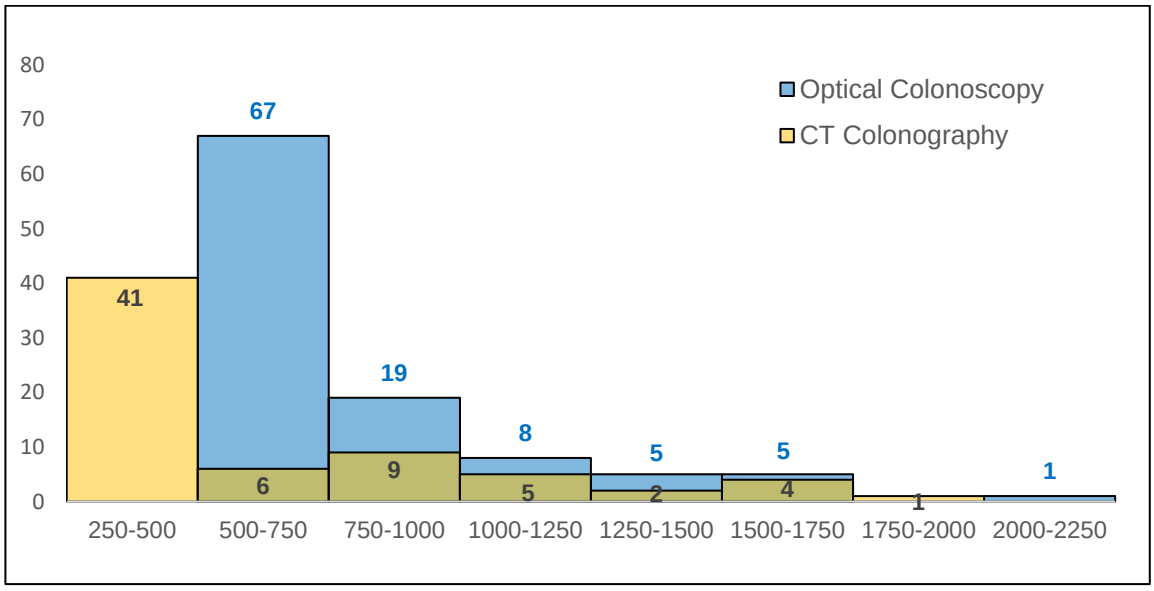


Figure 60. Overlapped histogram for the 6-month cost distribution (in £) for the OC (in blue) and the CTC group (in yellow).

Given the non-randomised study design, a second GLM analysis was performed to adjust for baseline characteristics that were statistically different between groups using a conservative threshold (p<0.10) (see Table 67). The initial unadjusted 6-month cost difference between groups (-£345) was robust and hardly affected by baseline differences between groups (-£370) (Table 66 and Table 67). In all analyses, the utilisation of CTC as the initial investigation for these cohort of participants was associated with statistically significant cost savings for the NHS.

Table 67. GLM analysis for the 6-month cost analysis variable (gamma function, identity link) adjusted using statistically significant differences (p<0.1) between groups at baseline.

6-month cost	**Coef.**	**Std. Err.**	**t**	**P>|t|**	**[95% CI]**	
Age	3.122	4.503	0.69	0.488	-5.703	11.947
Bowel complaints	29.335	32.070	0.91	0.360	-33.521	92.191
Comorbidities	-2.532	27.053	-0.09	0.925	-55.556	50.941
Utilities at baseline	-15.716	245.634	-0.06	0.949	-497.15	465.72
Group	-369.679	94.220	-3.92	0.000	-554.35	-185.01
Constant	1085.13	361.43	3.00	0.003	376.74	1793.51

(ii) Cost analysis including cancer treatment costs

A second cost analysis included the costs associated with the treatment of cancer (CRC or otherwise). The mean cost per participant increased to £1,537 and £1,423 in the OC and CTC group, with the cost difference of -£114 not being statistically significant (p=0.759) (Table 68).

Table 68. Mean (SD) cost per participant including costs associated with cancer treatment.

Mean total cost (SD)	OC group (n=105)	CTC group (n=68)	Unadjusted cost difference (CTC group – OC group, p-value)
6 months	£1,537 (£536)	£1,423 (£450)	- £114 p=0.759

5.3.5 Secondary Objectives

Outcome 1: 6-month cost-effectiveness of CTC as the first imaging modality compared to OC.

At baseline, participants in the OC group showed a trend of higher utility value (mean utility of 0.793 vs 0.832, p=0.097) using the EQ-5D-5L questionnaire. EQ-5D-5L data were collected at 3 and 6 months post-recruitment (Table 69). There were no statistically significant differences (p>0.05) between the groups at either 3 or 6 months in relation to the utility scores.

Table 69. Descriptive statistics for the utility variable at baseline, 3 and 6-month post-recruitment.

		N	Mean	Standard Deviation	p-value
Utility at baseline	OC	105	0.832	0.197	0.097
	CTC	68	0.793	0.173	
Utility at 3 months	OC	58	0.792	0.218	0.603
	CTC	35	0.802	0.171	
Utility at 6 months	OC	58	0.818	0.213	0.390
	CTC	44	0.793	0.217	

The mean cost per QALY at month 6 was estimated at -£69,080 (i.e. the intervention dominates, producing marginally more QALYs at a lower cost) (Equation 11).

Equation 11. Point estimate of the incremental cost per QALY at month 6.

$$ICER = \frac{Cost\ CTC - Cost\ OC}{QALY\ CTC - QALY\ OC} = \frac{-£345.3}{0.005} = -£69{,}080$$

Figure 61 illustrates the bootstrap analysis with 1,000 replicates, considering the 6-month cost per QALY. At month 6, the use of CTC as the initial investigation had a probability of 56.0% of being dominant and 0.0% of being dominated by the control group. The remaining 44.0% of bootstraps were in the cost-effectiveness analysis quadrants, i.e. the probability of being cost-effective depends on the overall system willingness-to-pay for each QALY. Assuming a £20,000 and £30,000 willingness-to-pay per QALY (thresholds typically considered by NICE), there was a 91.4% and 83.6% of CTC being cost-effective compared to OC at six months, respectively (Figure 62).

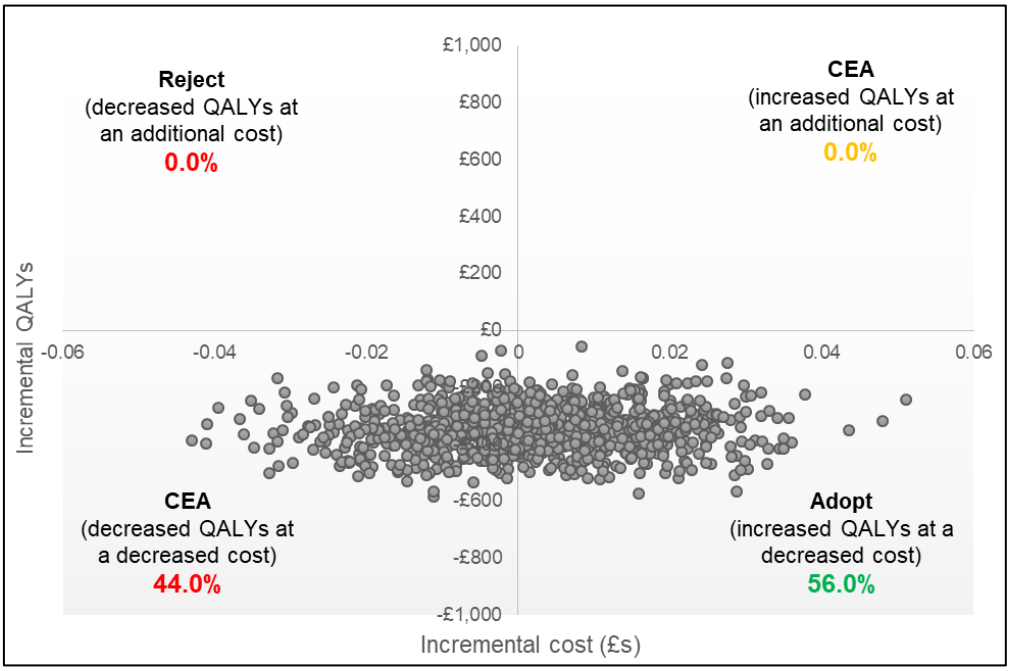


Figure 61. Cost-effectiveness plane associated with the 6-month cost per QALY analysis and probability associated each quadrant (bootstrap analysis with 1,000 replicates).

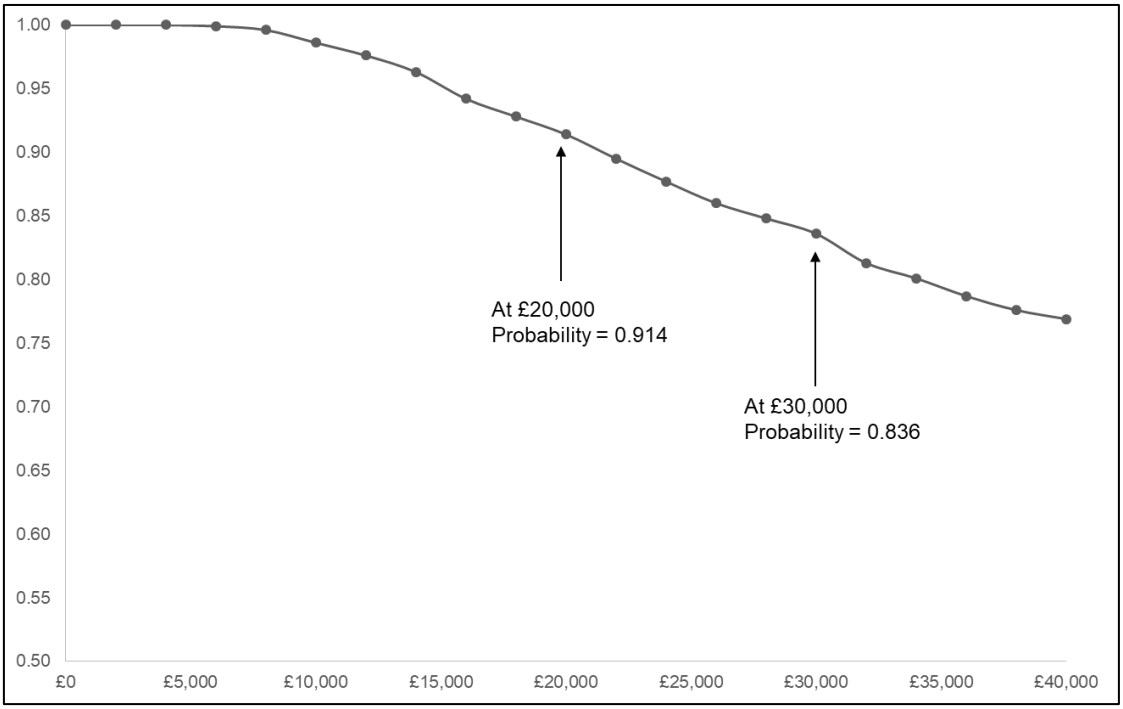


Figure 62. Cost-effectiveness acceptability curve for several thresholds of willingness-for-pay.

Objective 2: Cost per correctly diagnosed CRC using CTC as the initial investigation compared with OC.

Out of the 68 CTC scans, 22 invasive colonic investigations were subsequently performed, 11 OC and 11 flexible sigmoidoscopy scans. Of the 22 subsequent scans performed, there were no discrepancies in the diagnosis of medium/large polyps and CRC between the CTC and a subsequent invasive colonic test.

Objective 3: To estimate the incidence of extracolonic findings in the CTC group.

Out of the 68 participants in the CTC group, 17/68 (25%) had extracolonic findings, 3/17 (18%) of which warranted further assessment. The first participant had a mass in the omentum that led to a laparotomy with surgical excision, the second presented an inguinal hernia that led to a follow-up CT and an initial outpatient appointment at which no further action was deemed necessary and the third participant was diagnosed with a Bosniak type IIF left renal lesion that after a follow-up CT warranted no further action.

Objective 4: Likelihood ratio of CTC against colonoscopy in patients with initial positive CTC findings for medium to large polyps (≥6 mm) and CRC.

No false positive findings in the initial CTC test were found when compared with the reference OC / flexible sigmoidoscopy. Therefore, it was not possible to estimate the test's likelihood ratio as per the equation below.

$$\text{Likelihood ratio (LHR) for a positive CTC} = \frac{\text{True Positive}}{\text{False Positive}} = \frac{22/22}{0/22}$$

Objective 5: Time taken to reach a definitive diagnosis and first major treatment decision.

Access to care was measured using time from: (i) referral to test; and (ii) referral to report being available to the referrer as proxies for accessibility (Table 70). In terms of the time elapsed from primary care referral to diagnostic test, there were no differences between the two groups (p=0.698), with mean time (SD) of 14.6 (6.3) and 12.8 (7.3) days for the OC and CTC group, respectively. With regards to the time from referral to the report being available to the referrer, the mean (SD) time was 34.2 (18.0) and 27.9 (16.6) days for the OC and CTC group, respectively, with the 6.3 days' time difference being statistically significant (p=0.005). A detailed analysis of both variables showed that patients undergoing OC as the initial diagnostic test had a higher probability of having biopsies taken that ultimately led to a time delay in the report being available to the referrer.

Table 70. Mean (SD) time elapsed from (i) referral to test; and (ii) referral to report being available to the referrer organised by study group.

	OC group (n=105)	**CTC group (n=68)**	**p-value ***
Mean time elapsed in days (SD) between referral and the diagnostic test	14.6 (6.3)	12.8 (7.3)	0.698
Mean time elapsed in days (SD) between referral and the report being available to the referrer	34.2 (18.0)	27.9 (16.6)	0.005

** Non-normal distributions estimated using the Shapiro-Wilk test. Hence, Mann-Whitney tests were performed.*

Objective 6: Cancellation rates and incomplete (or suboptimal) bowel investigations.

OC and CTC had similar rates of incomplete or suboptimal bowel visualisation (p=0.802), with, respectively, 9 (8.6%) and 6 (8.8%) scans deemed not suitable and thus requiring a second test. Similarly, no differences on the day cancellation rates were found between the groups (p=0.459), with 1/105 (1.0%) cancelled OC (refused to undergo procedure on the day) and 2/68 (2.9%) cancelled CTC scans (patients could not tolerate the test). In these 2 CTC scans, one patient underwent a CTC scan with sedation whilst the second underwent an OC scan instead.

Objective 7: Patient satisfaction.

Patient satisfaction was assessed the morning after the diagnostic scan and asked the participant's opinion about: (i) the bowel preparation (pre-test); (i) the bowel test (test) and (iii) the morning after

the test (post-test). Appendix IV summarises the responses from participants in the OC group (n=69) and the CTC group (n=45). No statistical differences were detected between the two groups before, during or after the initial test (OC or CTC scan).

5.3.6 Sensitivity Analyses

Deterministic sensitivity analyses around the unit costs of CTC and OC scan were performed. The results from the sensitivity analyses are presented in Table 71.

(i) CTC scan. The unit cost per CTC scan was a key variable in the cost analyses, particularly in the CTC group. The £250 unit cost was based on NHS reference costs and considered both the time elapsed in the acquisition of the CT images as well as the subsequent radiologist reporting. This unit cost was varied - 25%/+25%, equivalent to £187.50/£312.50, leading to cost differences per participant at 6 months of, respectively, -£405 ($p<0.001$) and -£288 ($p<0.001$). This meant that the use of CTC led to cost savings in both deterministic scenarios.

(ii) OC scan. Similar to the previous sensitivity scenario, the unit cost per OC was considered as a second sensitivity scenario. Contrary to CTC, which unit cost was estimated based on a single unbundled tariff, the unit cost of OC scans varied according to the type of procedure (e.g. diagnostic colonoscopy, therapeutic colonoscopy, combined upper and lower endoscopic procedures – see Table 63 for further detail), leading to a minimum unit cost of £515 and a maximum of £911. All unit cost were varied -25%/+25%, leading to cost differences per participant at 6 months of, respectively, -£168 ($p=0.012$) and -£524 ($p<0.001$). This meant that the use of CTC led again to cost savings under both deterministic scenarios.

Table 71. Sensitivity analyses scenarios considered and respective impact in the cost analyses at 6 months post-recruitment.

	Mean cost difference		**p-value**	
Base case scenario	- £345		<0.001	
(i) **CTC unit cost:** -25% / +25% variation	-£405	-£288	<0.001	<0.001
(ii) **OC unit cost:** -25% / +25% variation	-£168	-£524	0.012	<0.001

5.3.7 Summary of Results

Participants in the CTC group were older, with a mean age of 69.2 years old compared to 61.3 in the OC group, with a greater disease burden, a higher number of comorbidities / active health problems (4.0 vs 2.9 in OC group) and lower utility values at baseline (0.793 in the CTC group vs 0.832 in the OC group). These imbalances were considered in the adjustment of the primary outcome. Five cancers (4.8%) were diagnosed during the 12-month follow-up period in the OC group compared to six cancers (8.8%) in the CTC group, with this difference not being statistically significant. Out of the 68 participants in the CTC group, 17 (25%) had extracolonic findings, 3/17 (18%) of which warranted further assessment.

With regards to the primary outcome, the utilisation of CTC as the first-line colonic investigation for patients with low to intermediate risk of CRC led to a lower mean cost per participant at 6 months compared to the use of OC. Based on unadjusted GLM distributions, the mean cost difference per participant between groups at 6 months was - £345 (CI 95% CI: -£501 to -£190). When taking into consideration baseline imbalances, the mean cost difference between groups marginally increased and remained statistically significant [-£370, CI 95% CI: -£554 to -£185)]. An additional cost analysis included the costs associated with cancer treatment (CRC or other cancers). The mean cost per participant increased, respectively, to £1,537 and £1,423 in the OC and CTC group, with the cost difference of -£114 not being statistically significant (p=0.759).

Assuming a £20,000 and £30,000 willingness-to-pay per QALY (thresholds typically considered by NICE), there was a 91.4% and 83.6% probability of CTC being cost-effective compared to OC at six months, respectively.

5.4 Discussion

5.4.1 Overview

This section discusses the clinical and economic findings from the colon study. Furthermore, strengths and limitations are discussed, along with the potential implications for research and clinical practice in the management of patients with suspected CRC.

5.4.2 Aims and Objectives

The study's primary objective considered the 6-month cost implications of using CTC as the initial colonic investigation compared to OC from a healthcare payer perspective. This decision was based on the absence of high-quality economic evidence in the selected subpopulation of patients (low to intermediate risk of CRC). Secondary objectives considered cost-effectiveness and cost-utility analyses and other dimensions of analysis, from accessibility to care, patient satisfaction and diagnostic accuracy of both clinical pathways. These outcomes aimed to evaluate the real-world implications of using CTC compared to OC across a wide range of dimensions of care in the NHS.

5.4.3 Key Findings

Primary outcome:

6-month cost analysis

The study showed that the use of CTC produced cost savings to the NHS at 6 months post-recruitment, with an unadjusted mean cost difference per participant of -£345 (95% CI: -£501 to -£190, p<0.001). When adjusting for baseline imbalances between both groups, the mean cost difference per participant increased to -£370 (95% CI: -£554 to -£185, p<0.001). The cost differences between the unadjusted and adjusted analyses were due to the fact that participants in the CTC group were, on average, 8 years older and presented with a higher disease burden.

The cost difference between the CTC and OC groups were multifactorial, primarily driven by two factors: (i) the low incidence of colon findings; and (ii) the lower unit cost (£250) of CTC compared to OC (£515 to £760). The study's underlying hypothesis was that the use of CTC as the initial investigation for patients with low to intermediate risk of CRC would avoid the need for invasive diagnostic scans (such as OC) that are more expensive to the overall healthcare system. Atkin et al. (2013) reported that only 8% of patients had additional colonic investigations following OC, compared to 30% after CTC. These findings were corroborated by our study, with 11.4% and 32.4% participants having a subsequent colonic investigation in the OC and CTC group, respectively. These conversion rates drove the overall cost for the healthcare system. In essence, more than the tests' accuracy, that are considered to be equivalent, the presence of positive or negative findings was determined by the incidence of medium to large polyps or CRC in the population included in the study (overall 4% CRC incidence rate). This was indeed the reason to deliberately target the use of CTC to patients with low to intermediate risk of CRC. These patients are less likely to require a biopsy during an invasive colonic procedure, thus maximising CTC's ability to rule-out any major colonic finding. Out of the 68 participants in the CTC, 46 (68%) were discharged based on the CTC results alone. If patients with red flags (e.g. rectal bleeding, anaemia) were to be included, a higher proportion of patients would be expected to require an OC, thus reducing the cost savings or even increasing total costs to the healthcare payer.

The cost difference per patient estimated in this study was compared against the three NHS studies considered in the economic review (Table 72). In all four studies, the utilisation of CTC led to cost savings compared to OC. However, our study presented significantly higher cost savings compared to the other studies, particularly the RCT that showed no statistically significant cost differences (Halligan et al. 2015). Compared to the RCT, the unit cost per CTC group was similar (£645 vs £674) but there was a difference of £250 in the OC group (£991 in our study vs £739 in the RCT). The latter seemed to be due to the combined interaction of three factors. First, the incidence of CRC was higher in the RCT compared to our study (5.5% vs 4.3%), leading to increased costs in both groups of the RCT. Second, the unit cost of both CTC and OC used in the RCT were significantly lower than the ones considered in our study (mean unit cost of OC ranges from £330 to £450 in the RCT compared to £515 to £911 in our study and the mean unit cost of CTC of £160 in the RCT compared to £250 in the colon study). This ultimately drove down the unit costs per

participant in the RCT compared to our study. Third, only applicable to the CTC group, costs in our study included the management of extracolonic findings whilst, in the RCT (Halligan et al. 2015) and the other two studies (Ratnasingham et al. 2017; Gomes et al. 2013) excluded these costs.

Table 72. Comparative analysis of the cost findings from the colon study and three NHS cost analyses retrieved from literature review.

	Colon study (current study)	Ratnasingham et al. (2017)	Atkin et al. (2013) and Halligan et al. (2015)	Gomes et al. (2013)
Study design	Observational prospective study	Retrospective study	Randomised Controlled Trial (SIGGAR trial)	Economic Modelling (Markov)
OC group: mean cost per participant	£991	£737	£739	£583
CTC group: mean cost per participant	£645	£550	£674	£467
Mean cost difference per participant	-£345	-£187	-£65	-£116

As a corollary, the colon study produced statistically significant cost savings for patients with low to intermediate risk of CRC, contributing to the body of evidence concerning the utilisation of CTC as a direct alternative of OC in a major NHS Trust.

Secondary outcomes:

Cost-utility analysis

In addition to the 6-month cost analysis, we conducted a cost-utility analysis, comparing the incremental costs and QALYs as the measure of effect (NICE 2011b; 2013). The use of CTC was found to be cost effective compared to OC, achieving more QALYs at a lower cost. Considering NICE's traditional willingness-to-pay thresholds of £20,000 to £30,000 per QALY, CTC was found to have a probability of 91% and 84% of being cost-effective, respectively. These findings contrast to the 60% probability of cost-effectiveness estimated by Gomes et al. (2013). The very high likelihood of CTC being cost-effective was mainly driven by cost savings as 44% of replicates CTC led to lower QALYs compared to OC. Hence, the more the healthcare system is willing to pay, the less likely the intervention with CTC is to be cost-effective (91% vs 84% at £20,000 to £30,000 per QALY, respectively). However, these findings were potentially affected by the attrition rates associated with the 6-month EQ-5D-5L questionnaire (45% and 35% in the OC and CTC group, respectively). The utilisation of multiple imputation methods for missing utility data at months 3 and 6 did not affect the cost-utility analysis, with the intervention with MRI presenting similar probabilities of being cost-effective (92% vs 86% at £20,000 to £30,000 per QALY, respectively).

Likelihood ratio for a positive CTC

Previous clinical studies have established the non-inferiority of CTC compared to OC in the diagnosis of medium to large polyps (≥5mm) and CRC (NICE 2018a). This finding was corroborated by our study as all positive findings in the CTC were confirmed (i.e. true positives) based on a subsequent invasive test (either OC or flexible sigmoidoscopy). No CRC was diagnosed up to 12 months post-recruitment in participants with negative findings in the initial CTC. Although reassuring, this was a study limitation as patients with negative findings in the initial CTC did not undergo an invasive test as per routine clinical practice. These findings were compared with the SIGGAR trial, which considered a longer follow-up period time (3 years vs 1 year), and both studies corroborated the non-inferiority of CTC compared to OC in the management of CRC.

Extracolonic findings

CTC enables the visualisation of extracolonic findings that may not relate to the presenting clinical condition but may be of clinical significance, e.g. abdominal aortic aneurysms which are potentially life threatening if untreated (Pooler, Kim, and Pickhardt 2017). Out of the 68 participants in the CTC group, 17 (25%) had an incidental extracolonic finding. Out of these 17 participants, 14 (82%) did not warrant any follow-up, with the remaining 3 (18%) requiring follow-up with diagnostic CT follow-up and, in one case, a surgical excision of a mass. Evidence from the SIGGAR trial showed a much higher proportion of CTC scans with at least one extracolonic finding (59%, 1039/1748), with 14% (149/1039) of these patients undergoing subsequent procedures to further investigate and/or treat the extracolonic findings. Although the proportion of extracolonic findings reported in the SIGGAR trial was much higher compared to the colon study (59% vs 25%), the proportion of patients with extracolonic findings requiring further investigations and/or treatment was similar in both studies (14% vs 18%). It was hypothesised that, given the randomised controlled design of the SIGGAR trial, radiologists reported clinical findings that would not otherwise be reported in real-world practice (CTC scans in the colon study were performed as part of standard care).

In addition, the utilisation of CTC led to the diagnosis of two patients with extracolonic cancer, a metastatic stomach cancer and a metastatic prostate cancer (2.9%). These extracolonic findings were not believed to be incidental but rather related to the presenting bowel complaint. The incidence of extracolonic cancers was similar with the evidence from previous studies [2.7% based on an extensive study performed by Xiong et al. (2005)]

Access to care

The use of CTC as the initial investigation led to the results being available to the referrer, on mean, 6.3 days earlier compared to the OC group (p=0.005). The longer waiting time in the OC group was mainly due to the diagnostic workload associated with biopsies taken during the OC test rather than the date of the OC test itself. As part of the OC procedure, biopsies of potential lesions (polyps or CRC) were often taken and the time it took to complete the pathological assessment of those tissues seemed to be the major contributing factor for the difference in access to care between the two groups. However, this time difference was not deemed to be clinical relevant.

On the day cancellation and incomplete scans

OC and CTC had similar rates of incomplete or suboptimal bowel visualisation in the colon study, with, respectively, 9 (8.6%) and 6 scans (8.8%) deemed unsuitable and thus requiring a second test. The SIGGAR trial reported a higher rate of incomplete OC scans (11%) and a lower rate of incomplete CTC scans (5.3%) (Halligan et al. 2015).

No differences on the day cancellation rates were found between the groups, with 1.0% and 2.9% cancellation rates among patients undergoing OC and CTC scans, respectively. Given its operational impact, particular emphasis was given to reduce 'on the day cancellation' and 'did not attend' rates as discussed in further detail in Chapter 7.

Patient satisfaction

Patient satisfaction was compared using a non-validated questionnaire to be completed the morning after the diagnostic scan (CTC or OC) and evaluated the participant's opinion about: (i) the bowel preparation (pre-test); (ii) the bowel test (test); and (iii) the morning after the test (post-test). No statistically significant differences between groups were found between the two groups at any of these three phases. These findings were somewhat unexpected, particularly concerning the bowel test itself and the experience on the morning after the bowel test (post-test). It was assumed that, given the non-invasive nature of CTC, participants in this group might report shorter recovery times, lower levels of discomfort or intrusiveness or improved experience in comparison with their expectations. The latter could be due to the small number of questionnaires completed [n=70 (67%) for the OC and n=45 (66%) for the CTC group]. For this reason, and bearing in mind the increase in the uptake of CTC (see Chapter 7 for further detail), it was recommended to the gastroenterology and radiology departments to continue to use patient experience questionnaires. This should provide definitive evidence to evaluate any differences in patient satisfaction between the two tests.

5.4.4 Strengths, limitations and implications

Strengths

The estimates of NHS resource use data were primarily based on comprehensive and complete data retrieved from hospital-based databases that captured both the acute and elective elements of the pathway associated with the management of patients with low to intermediate risk of CRC. These data were supplemented by both primary care utilisation data, collected from each participant's GP, and self-reported participant data. The aim was to guarantee that any colon-related NHS event was costed regardless of the healthcare provider and its location. Moreover, the inclusion of costs associated with the management of extracolonic findings were also included. This contrasted with the majority of the economic literature and constituted another strength of this study. In addition, the prospective collection of healthcare utilisation and self-perceived quality of life data, the conduction of economic evaluation analyses (rather than cost analyses only) and the use of different dimensions of analysis (efficiency, quality of care, access to care and patient satisfaction) were other key factors that contributed to the overall strength of the study.

Limitations

There were however some limitations to this study. First, this was a single-centre study with participants recruited from one central hospital in London. A multi-centre study would be necessary to explore the generalisability of the results. Second, as with any observational study, no randomisation between groups was performed. An adjusted GLM was performed to mitigate the potential impact of the study observational design on the primary outcome. Third, the study sample might not be representative of all patients with low to intermediate risk of CRC. To mitigate this potential impact, clear and detailed inclusion and exclusion criteria were used. Fourth, participants with negative CTC findings could not be definitively ruled-out as potential false negatives as they did not undergo an OC as per routine clinical practice. The supplementary use of cancer databases at one year follow-up minimised this limitation. Fifth, there were considerable follow-up attrition rates potentially affecting secondary outcomes such as cost-utility analyses or patient satisfaction. This issue could be minimised with closer cooperation between the clinical and the research teams in the proactive follow-up of participants. Lastly, long-term clinical and economic implications of extracolonic findings found in the CTC group were not evaluated as part of this study.

Implications for Further Research

The utilisation of CTC as a replacement for OC as the first-line colonic investigation should be evaluated across different population subgroups, care providers and healthcare settings. Furthermore, any future studies should evaluate the impact of both CTC and OC scans on patient satisfaction and test acceptability. Finally, the long-term impact of extracolonic findings in CTC scans should be further evaluated in larger sample sizes and modelled beyond the study's time horizon to estimate the cost-effectiveness of CTC compared to OC.

Implications for Policy and Clinical Practice

The utilisation of CTC as a first-line replacement test for OC in the diagnosis of symptomatic patients with change in bowel habits as the presenting complaint should be considered. CTC is a safe, cheaper and more accessible alternative to OC in the diagnosis of symptomatic patients with low to intermediate risk of CRC. A quarter of patients undergoing CTC scans presented extracolonic findings, most of which did not warrant any further investigation and/or treatment. However, clinical literature from the SIGGAR trial reported that more than half of patients had at least one extracolonic finding. Given its potential resource implications in the utilisation of CTC as a first-line diagnostic tool, report standardisation of extracolonic findings in CTC scans should be considered, along with clear diagnostic and/or therapeutic clinical management pathways.

5.5 Conclusion

This study found that the use of CTC instead of OC as the initial investigation scan for patients referred from primary care with low to intermediate risk led to significantly lower costs to the NHS at 6 months post-recruitment. Furthermore, the use of CTC had a high probability of being cost-

effective at month 6 using NICE's willingness-to-pay thresholds. The use of CTC also improved access to care and no difference in patient satisfaction was noted.

In summary, the use of CTC for patients with low to intermediate risk of CRC led to savings for the NHS whilst enabling the release of vital OC capacity to enhance access to care for patients more likely to benefit from an invasive procedure.

Chapter 6. Trial-based to model-based economic evaluations

6.1 Chapter overview

This chapter summarises the utilisation of economic data from diverse study types. First, a literature review discusses the strengths and weaknesses of using economic data from multiple study designs, with particular focus on data derived from interventional studies [e.g. randomised controlled trials (RCTs)], real-world evidence (e.g. observational studies) and decision analytical modelling studies. Second, the potential implications of using different study types in the context of the TOHETI programme are discussed. For this purpose, the *a priori* economic modelling conducted as part of the TOHETI bid is compared to the actual findings from the RCT and two observational studies (findings already detailed in chapters 3, 4 and 5). Differences in economic evidence between trial-based and model-based analyses are presented, together with a discussion of potential implications to policy makers and NHS clinical practice.

6.2 Literature review: trial-based to model-based economic evaluations

The introduction of new medical technologies has been recognised as one of the main drivers associated with a growth in healthcare costs, estimated to be responsible for 40-50% of annual cost increases (Clemens, 2017). However, a review by Sorenson, Drummond, and Khan (2013) analysed 86 studies and found conflicting evidence, with some technologies leading to increased costs (e.g. cancer drugs, invasive devices), while others were cost-neutral or even cost saving. The introduction of such technologies should therefore be preceded by an economic evaluation based on high-quality evidence, focusing not only on costs but also on respective effects.

One fundamental discussion concerns the data and study design(s) on which any economic evaluation is based on. Drummond (1996) first discussed the need to find a balance between ideal clinical-trial based economic evaluations, with high internal validity but low external validity, and economic models, which may be more relevant to the decision maker. Historically, information collected in RCTs represented the dominant paradigm in economic evaluation. A study by Sculpher et al. (2006) reported that approximately 30% of economic evaluation studies on the NHS Economic Evaluation Database were derived from single RCTs. However, deriving pivotal data, e.g. resource use, from a single-centre study could bias economic findings due to potential lack of generalisability. In 2007, ISPOR published a recommendation supporting the use of real-world evidence (RWE), i.e. clinical, economic and patient-reported data based on studies that were not RCTs such as cohort and case-control studies (Garrison et al. 2007). Figure 63 illustrates the different levels of evidence typically produced during the timeline of the approval of new technologies.

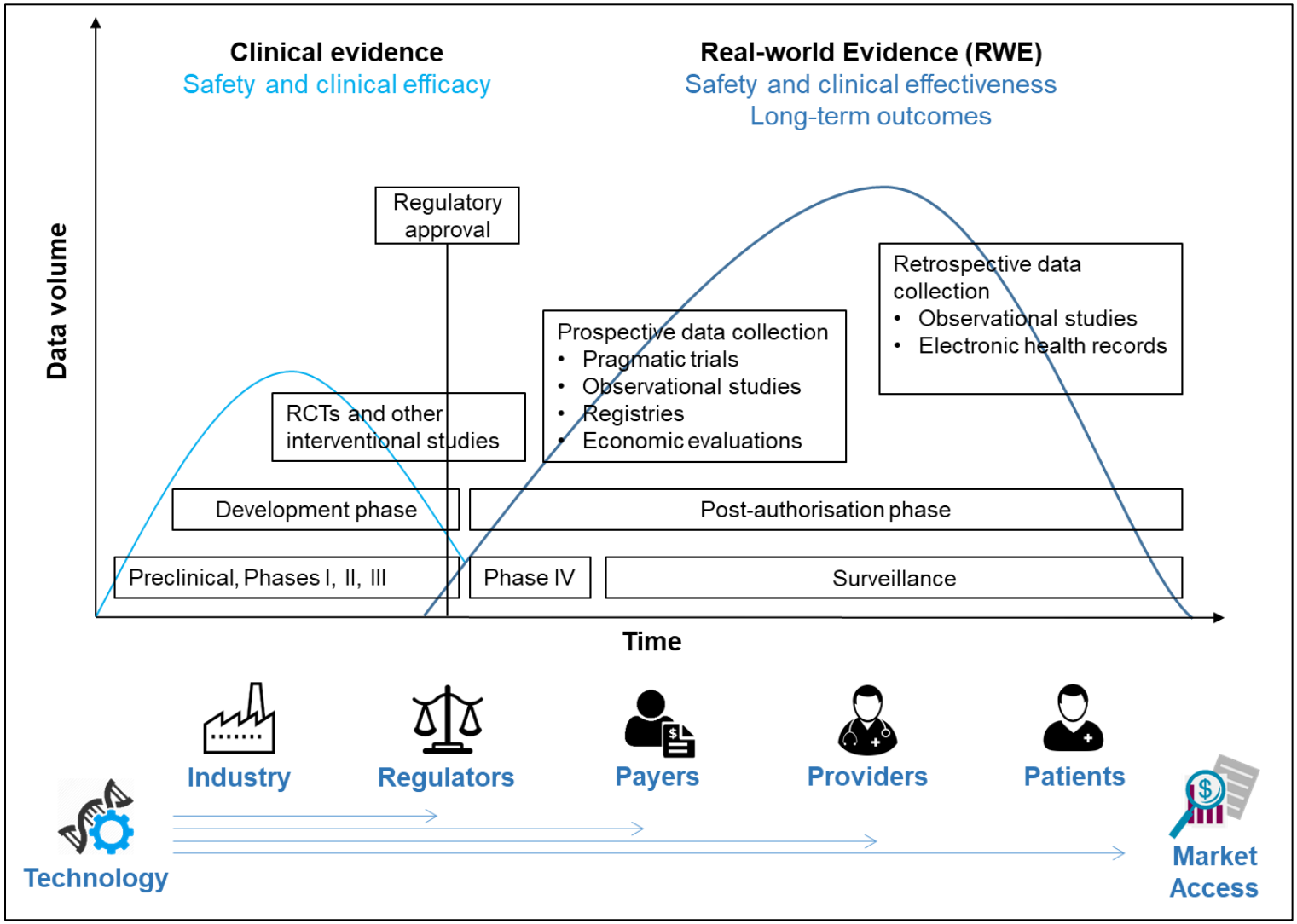


Figure 63. Schematic illustration of the type of study design and respective data volume generated (graph lines in shades of blue) across the technology decision approval process (Katkade, Sanders, and Zou, 2018).

RCTs are mainly conducted to determine an intervention's efficacy prior to regulatory approval. Contrastingly, a higher volume of RWE is usually generated following regulatory approval, ranging from pragmatic RCTs to observational studies and economic models.

A literature review capturing the strengths and weaknesses of trial-based or model-based economic evaluations is detailed in Appendix V. In summary, trial-based economic evaluations are based on observed data but typically do not compare all alternatives, present truncated follow-up periods, do not assess final outcomes or lack generalisable findings. Model-based evaluations address these issues but present potential biases of their own. Some of these biases include: the incorporation of inadequate clinical data; unknown effect of the intervention; and inadequate statistical methods to address biases from different types of data (e.g. observational studies). The use of pragmatic RCTs is increasingly recognised as the most relevant study design to evaluate the cost-effectiveness of interventions in the context of real-world clinical practice. Pragmatic RCTs aim to increase the generalisability of findings but consideration needs to be given to a potential decrease in internal validity, affecting the trial's ability to make causal inferences (Calvert, Wood, and Freemantle, 2011).

Data derived from multiple study designs should ideally be used to assist the decision maker in the context of specific interventions. The type of data informing a technology economic evaluation depends on: (i) the technology itself, particularly its value proposition and likelihood to impact routine clinical practice; and (ii) the level of existing clinical and economic evidence. The more disruptive the technology, the less likely it is to have published clinical and economic evidence, such as economic evaluation conducted alongside an RCT or a pragmatic RCT. In contrast, if the technology's anticipated benefits, both in terms of costs and effects, cannot be estimated based on RCTs or RWE studies, economic modelling provides a solid framework for evidence synthesis and decision modelling on which to base a funding decision (Katkade, Sanders, and Zou, 2018). Rather than providing a definitive answer as to which it is better, one of the aims of this chapter is to provide the reader with an understanding that the choice of method is dependent on the context of the technology being evaluated.

The following subsections compare the economic findings from three cases studies based on real-world studies (one pragmatic RCT and two observational cohort studies, chapter 3 to 5) conducted in Phase 3 of the TOHETI programme with the *a priori* economic models built in Phase 2.

6.3 Methods

Phase 2 of the TOHETI programme included the funding proposal where decision-analytical models were presented to estimate the potential impact of the proposed interventions. Literature reviews were conducted to synthesise the clinical and economic evidence. This process assessed whether existing evidence had already evaluated the proposed intervention and, if so, summarised the evidence. Furthermore, the review of clinical evidence targeted studies where diagnostic accuracy (sensitivity, specificity, positive and negative predictive value) was evaluated directly or indirectly against the

standard care. Special consideration was given to systematic literature reviews and meta-analyses. With regards to economic evidence, any cost analysis or health economic evaluation (cost-effectiveness, cost-utility, cost-consequence or cost-benefit analyses) evaluating the intervention (or similar interventions) directly or indirectly against a comparator were included. Further detail for each literature review is presented under each clinical condition in the following subsections (6.4.1, 6.5.1. and 6.6.1). This evidence was essential to populate the *a priori* models built by the PhD student.

Subsequently, decision-tree models were developed in 2014-15 for each of the three clinical interventions considered. A decision tree is a tree-like decision support tool in which each node represents a "test" (e.g. patient undergoing MRI scan), each branch represents the outcome of such test (e.g. normal or abnormal findings in the MRI), and each leaf node represents a class label (e.g. patient with true positive findings). Given that the model probabilities are based on a one-off event (i.e. a diagnostic scan), decision tree models are used very often in diagnostic modelling. Each model compared the proposed intervention with the existing standard care. Although specific aspects of each economic model are presented in the following sections, all three models shared four general features. First, the main aim of the economic models was to estimate the cost implications of the different interventions from a healthcare payer perspective. A mean cost per participant was estimated for all three interventions. Second, the time horizon of the models was consistent with the follow-up period for each clinical condition. Third, the decision tree was structured to reflect the accuracy levels (sensitivity and specificity) of the proposed imaging interventions. Sensitivity refers to a test's ability to rule-in a clinical condition in patients with such condition, e.g. ability to identify a scaphoid fracture in patients with a scaphoid fracture (true positive). Specificity refers to a test's ability to rule-out a clinical condition in healthy patients, e.g. ability to exclude a scaphoid fracture in patients with no actual scaphoid fracture (true negative). Diagnostic tests do not have perfect accuracy in real-world clinical practice, potentially leading to false negative (e.g. patients with negative test results despite having a scaphoid fracture) and false positive findings (e.g. patients with positive test results despite having no scaphoid fracture). The different probabilities associated with these scenarios were considered in the decision tree models using sensitivity and specificity values retrieved from literature. Fourth, the overall incidence or prevalence of the condition being evaluated was included in the model. This was a key parameter given that all probability nodes will be affected by the proportion of patients with or without the actual condition. The more prevalent the condition is, the higher the proportion of patients with positive findings in the diagnostic test.

6.4 Suspected scaphoid fracture

6.4.1 Summary of literature review

This subsection summarises the key evidence used to model the impact of different imaging modalities (particularly MRI) in the management of patients with suspected scaphoid fracture. These key variables were the incidence of scaphoid fractures, and the accuracy levels of the different imaging modalities used in the management of suspected scaphoid fractures (summarised on Appendix VI). Other less

advanced imaging modalities, like bone scintigraphy or ultrasound, were not included in this literature review.

The incidence of scaphoid fracture was highly variable among patients recruited to the identified studies. A systematic review by Yin et al. (2010) showed a minimum incidence value of 5% and a maximum of 50%. However, these values derived from different reference tests and from multiple clinical pathways at different hospitals, partly explaining the registered variation. Taking this uncertainty into consideration, clinical evidence suggested a scaphoid fractures incidence value between 10 and 20% from the overall number of suspected scaphoid fractures (Yin et al. 2010). This is similar to the 10% and 16% incidence value reported, respectively, for patients with wrist injuries and patients with clinical suspicion of scaphoid fracture and normal initial radiography (Ring and Lozano-Calderón 2008, Mallee et al. 2011). For the purpose of the economic model a 10% baseline value was assumed and subsequently varied in deterministic sensitivity analyses.

Clinical literature supported the added diagnostic value of both CT and MRI compared to conventional radiography. Particularly, CT and MRI have a very high specificity value in the diagnosis of scaphoid fractures, i.e. are able to exclude scaphoid fracture among patients with no scaphoid fractures. High-quality clinical evidence from systematic reviews (e.g. Yin et al. 2010) supported MRI as the imaging modality with the highest accuracy level in the diagnosis of scaphoid fractures and, for this reason, it was included in the model. Sensitivity and specificity values of, respectively, 97.7% and 99.8% were included in the model based on the meta-analysis performed by Yin et al. (2010). Table 73 summarises the baseline incidence and sensitivity/specificity probabilities used in the decision tree model.

Table 73. Incidence, sensitivity and specificity parameters included in the economic model.

Parameter	Value	Source / Description
Estimated incidence of scaphoid fractures among patients with suspected scaphoid fracture presenting to the ED.	10%	Assumption. Sensitivity analysis was performed using literature.
Radiographs sensitivity and specificity in the diagnosis of scaphoid fractures at presentation.	Sens.= 64% Spec.= 90%	Nguyen et al. (2008).
MRI sensitivity and specificity in the diagnosis of scaphoid fractures.	Sens.= 97.7% Spec.= 99.8%	Yin et al. (2010)

6.4.2 Economic modelling

A *de novo* decision tree model with a time horizon of 3 months, based on a healthcare payer perspective, was developed to estimate the cost implications of using immediate MRI as an add-on test in the management of suspected scaphoid fractures compared with the standard care, which relied on the use of radiographs only in the ED. Figure 64 illustrates the model associated with the control group, i.e. current standard care, whilst Figure 65 shows the model structure associated with the intervention group, i.e. the provision of wrist MRI in the ED.

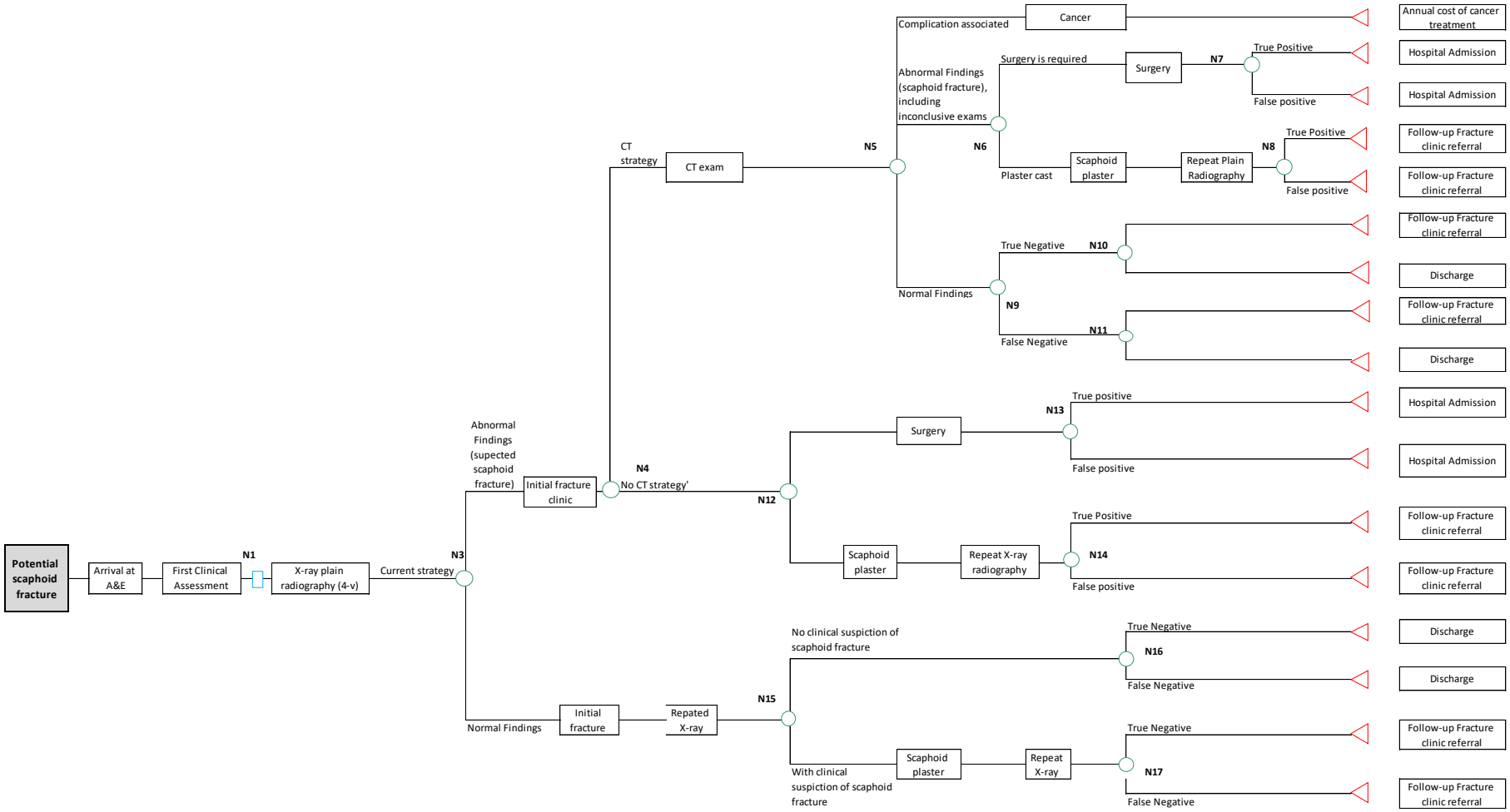


Figure 64. Short-term model for the control group (standard care) in the scaphoid model.

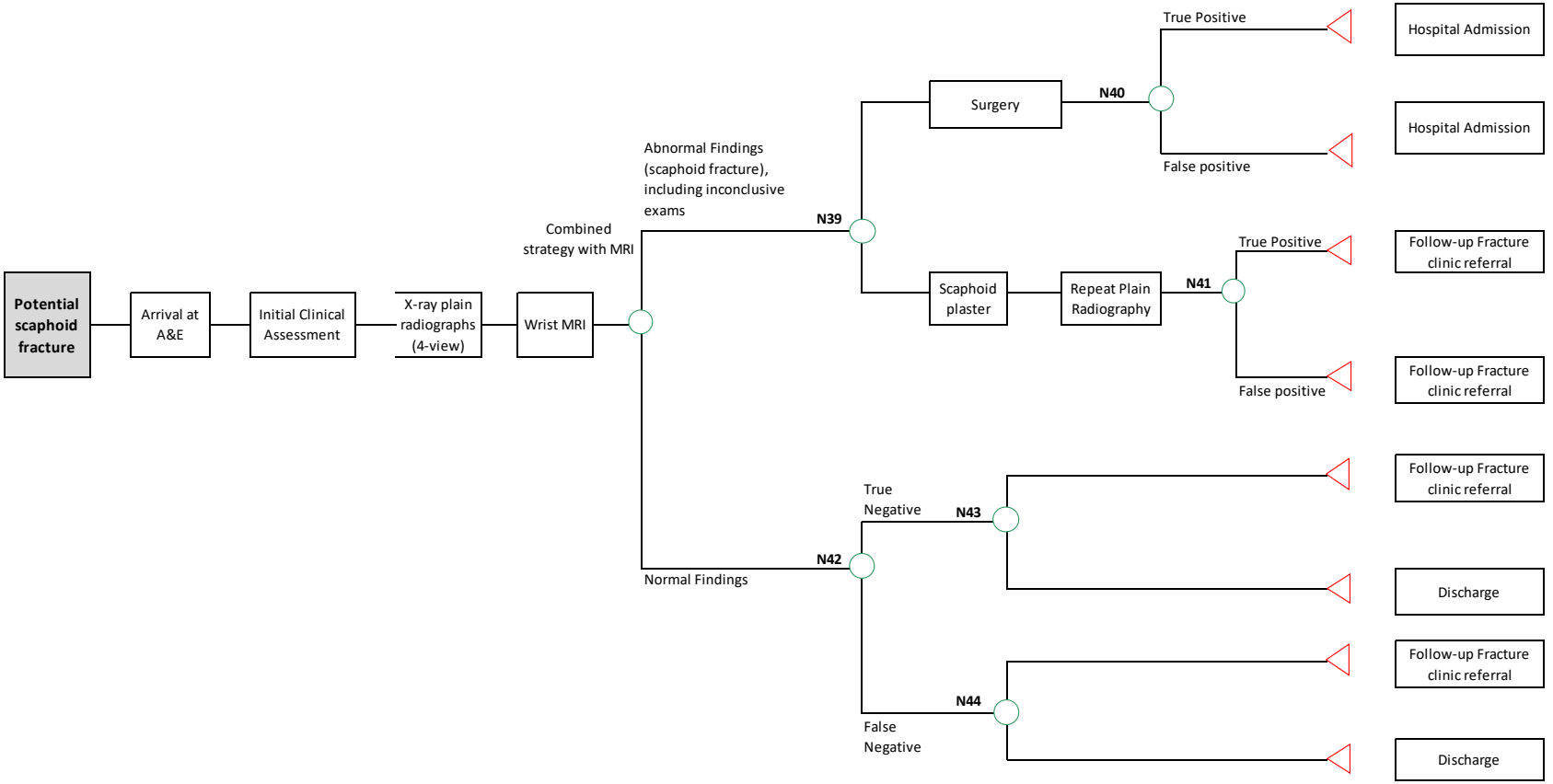


Figure 65. Short-term model for the intervention group (MRI group) in the scaphoid model.

The standard care group (Figure 64) comprised the combined use of two imaging modalities. First 4-view wrist radiographs were performed on presentation to the ED. If positive, a CT scan might be performed on a follow-up outpatient visit. This was usually carried out at an initial fracture clinic visit, a week to ten days post the initial ED presentation. Based on the scaphoid fracture incidence and the radiographs accuracy levels, 15.4% of findings were estimated to be positive. Applying these estimates of incidence and accuracy to an annual cohort of 679 patients attending the ED at GSTT, 105 patients would present with abnormal findings in the initial 4-view radiography performed at ED. Out of these 105 patients, 44 would be expected to be true positives (i.e. patients with scaphoid fractures and positive findings on the initial radiographs) and 61 to be false positives (i.e. patients with no scaphoid fracture despite positive findings in the initial radiographs).

The intervention group considered a combined strategy, with initial wrist radiographs and a subsequent MRI scan (Figure 65). The use of MRI was used as an add-on test to standard care following the initial wrist radiographs. All participants were then grouped into four logical branches as per the MRI findings: true positive, false positive, true negative and false negative. Patients with positive findings underwent multiple fracture clinic appointments and it was assumed that 5% eventually progress to surgery. Participants with true negative findings were discharged, but, based on clinical opinion, it was assumed that 25% of participants would still require follow-up due to ongoing pain. Participants with false negative findings were assumed to incur additional costs. The rationale was that patients discharged with scaphoid fractures would eventually re-present to the ED or to an outpatient appointment for further follow-up. Hence, in order for the model to consider this increase in utilisation, a cost of £430 was added (equivalent to cost of a repeated ED attendance, fracture clinic follow-up and a wrist MRI scan).

Table 74 lists the unit costs and branch probabilities included in the economic model. Unit costs were updated using the HCHS index (Curtis, 2012).

Table 74. Unit costs and other probabilities included in the economic model.

Parameter	Value	Source / Description
Percentage of patients with positive findings in the initial radiography that undergo a CT scan (control group)	80%	Assumption based on clinical experts (orthopaedic consultant).
Percentage of patients that undergo surgery after positive imaging findings (both groups)	5.0%	
Percentage of patients with a negative MRI scan that require a subsequent follow-up visit (MRI group)	25%	Assumption based on clinical experts. Sensitivity analysis is to be performed using a range of values.
Initial clinical assessment at the ED	£144.70	Patel et al. (2013). Unit cost included a removable scaphoid plaster cast ('backslab'). Consistent with internal hospital data.
4-view wrist radiographs in the ED	£28.70	Internal hospital data. Also consistent with Patel et al. (2013).
Initial fracture clinic visit	£155	Internal hospital data. Also consistent with Patel et al. (2013).
Follow-up fracture clinic referral	£91	Internal hospital data. Patel et al. (2013) reported a higher cost of £150.6.
CT exam	£60.60	Internal hospital data.
Cost per wrist MRI scan	£101.60	Internal hospital data.
Scaphoid plaster	£25.20	Internal hospital data. Patel et al. (2013) reported a higher value of £42.7.
Serial plain radiograph	£23.00	Internal hospital data. Patel et al. (2013) reported a similar cost (£18.7).
Surgery cost	£3,763	Internal hospital data.
Added cost in true positive patients	£0	No added cost in true positive patients.
Added cost in false positive patients	£0	No added cost was considered due to a false positive (resource use costs already considered in the model).
Added cost in true negative patients	£0	No added cost in true negative patients.
Added cost in false negative patients	£430	Added cost of a repeated ED attendance, fracture clinic appointment and conventional wrist MRI scan.

Table 75 presents the mean cost per patient for each strategy. The mean cost per participant in the control group (standard care) and intervention group (MRI group) were, respectively, £470 and £361, yielding potential savings per participant of £109 or 24% compared to the mean cost per participant in the control group. An estimate of £100 savings per participant was used in the power calculation for the randomised clinical trial summarised in Chapter 3.

Table 75. Mean cost per participant from the base case scenario for the control and intervention group.

Strategy (control vs intervention group)	**Cost per patient**	**Potential savings**
Control group: standard care -> diagnostic pathway in the ED based on radiographs only	**£470**	**-**
Intervention group -> combined strategy of radiographs and MRI	**£361**	**£109**

Sensitivity analyses:

One-way deterministic sensitivity analyses were performed around two model parameters. The sensitivity scenarios along with their respective results are presented in Table 76. The incidence of actual scaphoid fractures was varied to 5% and 50% (base case value of 10%) as per evidence from the systematic review by Yin et al. (2010). For the MRI group, the probability of formal follow-up for patients with negative findings in the MRI was varied to 0% and 100% (baseline value of 25%).

Table 76**.** Mean cost per participant from the base case scenario for the control and intervention group (MRI group) based on two deterministic sensitivity analyses.

Parameter	**Strategy**	**Sensitivity analyses**		**Cost per patient**	
		Minimum	**Maximum**	**Minimum**	**Maximum**
1. Incidence of scaphoid fracture amongst patients with suspected scaphoid fracture presenting at ED	Control Group	5%	50%	£ 453	£ 605
	Intervention group			£ 337	£ 555
2. Percentage of patients with a negative MRI exam that require a subsequent follow-up clinic visit (applicable to the MRI group)	Control Group	0%	100%	£ 470	£ 470
	Intervention group			£ 344	£ 412

If a 5% scaphoid fracture incidence value was assumed, the difference in cost per patient for the control and intervention group increased to £116. If the maximum incidence value of 50% was assumed, the cost difference per patient for the control and intervention decreased to £50. Hence, the higher the incidence of actual scaphoid fractures, the smaller the cost savings associated with the use of MRI. The latter is due to the fact that the added value of MRI resides in its ability to rule-out a fracture and decrease follow-up costs associated with these patients.

The probability of patients being discharged following a negative MRI was varied to 0% and 100%, given the uncertainty associated with this parameter. The higher the proportion of these patients being followed-up, the lower the cost savings associated with the use of immediate MRI. However, even if all patients were followed-up, the intervention was still associated with cost savings of £58 per participant.

6.4.3 Comparative analysis: real-world versus model results

Results from the RCT conducted as part of this thesis are presented in Chapter 3. The SMaRT trial was designed as a pragmatic RCT, taking into account different design and analytical considerations for trial–based health economics study (Marshall and Hux 2009). The intervention led to mean cost savings per participant of £174 and £266 at month 3 and 6 post-recruitment, respectively (Table 77). Hence, real-world evidence showed that the actual cost difference per participants was in fact higher than anticipated via model-based evidence.

Table 77. Mean cost per participant from the economic model and RCT for the control and intervention (MRI) group.

	Economic model results (3 months)		**Real-world data (randomised clinical trial)**			
	Cost per patient	Potential savings	3 months	6 months	3 months	6 months
			Cost per patient	Potential savings	Cost per patient	Potential savings
Control: Standard care with the combined strategy of radiographs and CT	**£470**	**£109**	**£542**	**£174**	**£661**	**£266**
Intervention: combined strategy of radiographs and MRI	**£361**		**£368**		**£395**	

The higher cost difference between groups estimated in the pragmatic RCT and in the economic modelling was driven mainly by the higher observed mean cost per participant in the control group, estimated at £542 and £661 at month 3 and 6, respectively. This was higher than the predicted mean cost of £470 using economic modelling. Real-world evidence from the RCT exhibited four trends that increased the estimated cost of diagnosing suspected scaphoid fractures in the control group. First, there was a considerable utilisation of follow-up radiographic and advanced imaging scans. A mean of 1.05 radiographs were used whilst advanced imaging (either CT or MRI) represented a mean of 0.46 scans per participant. That meant that almost half of participants in the control group ended up undergoing advanced imaging. Second, a higher utilisation of fracture clinic appointments (initial and follow-up appointments) was recorded. Third, the economic model did not consider the costs associated with the management of the condition within primary care. Although marginal (less than 3% of the total costs), primary care costs contributed to increase the cost difference between groups as participants in the control group consumed more primary care healthcare resources (mean number of GP visits of 0.35 and 0.11 in the control and intervention group, respectively). Fourth, unit costs considered in the real-world evidence were based on the year 2017/18 compared to 2012/13 in the economic model. Costs from the economic model were updated using the HCHS index (Curtis and Burns 2018), leading to an increased cost difference between groups in the economic model (from £109 to £118) that was closer to the estimates from the real-world trial.

In the intervention group, the mean cost per participant was similar between the RCT (£368 and £395 at 3 and 6 months, respectively) and the economic model (£361). However, there were differences in the composition of costs in the two approaches. First, the prevalence of non-scaphoid fractures was not considered in the economic model (only scaphoid fractures were considered). A total of 22% (15/67) of participants in the MRI group had non-scaphoid fractures, a proportion of which (around 50%) required treatment with immobilisation with plaster cast and radiographic follow-up. This led to an increase in costs in the RCT, with the economic modelling constituting an oversimplification of the clinical practice. Second, a proportion of participants with negative findings in both the initial radiograph and MRI still had at least one fracture clinic appointment. This was due to the fact that even in the absence of a bone fracture, participants were still symptomatic for wrist pain. The economic model assumed that 25% of participants with no findings in the wrist MRI would require at least one fracture clinic appointment. This figure, based on clinical experts, considered uncertain given the absence of empirical data, was confirmed to be 23% (10/43 participants) in the RCT. Third, as with the control group, primary care related costs were considered only in the real-world study. This contributed to an increase of less than 1% of the mean cost per participant in the RCT. Fourth, all patients with positive findings for fracture were assumed in the economic modelling to be immobilised with a plaster cast. Real-world clinical practice showed that this assumption was not correct as some fractures (e.g. trabecular fractures or non-displaced fractures) were actually immobilised with a splint rather than a plaster cast. The cost of a plaster cast is higher than a splint and tends to require more follow-up visits. Figure 66 illustrates the changes included in the economic model to better reflect clinical practice.

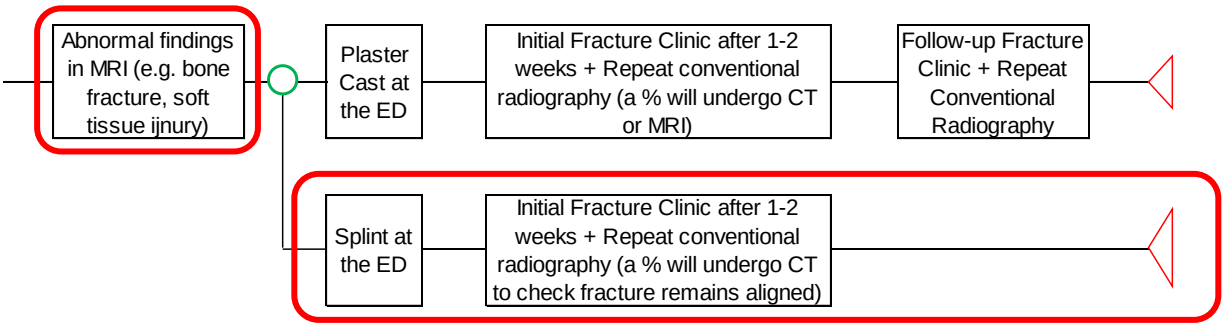


Figure 66. Changes to the scaphoid economic model based on trial-based data (highlighted inside the red squares).

In summary, the *a priori* economic model underestimated the cost impact of the intervention compared to data retrieved from the RCT. Real-world data showed that many of the assumptions underpinning the *a priori* model were inconsistent. These related mainly to both patient and clinicians behaviours. From the patient's point of view, the key driver to decrease utilisation of healthcare resources, and associated costs, derived from the MRI's ability to immediately rule-out a fracture (scaphoid or otherwise) or major soft injury (e.g. ligament rupture) during the acute episode of the pathway. Despite the MRI's reassurance effect, patients with no wrist injury might still remain symptomatic for a few weeks and consequently re-present to secondary care. This constituted a major parameter assumption for the economic model based on clinical expertise alone (no prior published evidence given the innovative nature of the intervention). The higher the proportion of patients re-presenting to secondary care, the less cost saving the intervention was. Another important finding from the real-world study, which was not considered in the economic model, were costs associated with the treatment of trabecular and aligned fractures. The economic modelling assumed that all participants with fractures, regardless of the location and nature, would be immobilised with a plaster cast. The real-world evidence showed that was not the case as non-scaphoid bone fractures (e.g. radial fracture) were not considered in the model and trabecular or aligned bone fractures were typically treated with a wrist splint rather than a plaster cast, decreasing the need for follow-up appointments and imaging. Given MRI's very high sensitivity, these types of fractures were detected mainly in the MRI group, ultimately decreasing the anticipated costs of treating these fractures.

6.4.4 Implications

As discussed in Chapter 2, there are five layers of uncertainty associated with the evaluation of medical imaging technologies. The disruptive nature of immediate MRI in the management of suspected fractures and consequent absence of empirical evidence from RCTs further increased the uncertainty around the cost implications to the NHS. This might ultimately lead to conflicting findings from model-based and trial-based studies, raising the question as to what level of evidence is required to assess such interventions. Gazelle et al. (2011) and the NICE Diagnostics Assessment Programme (Crabb,

2011) identified the need to obtain high-quality evidence about the introduction of technologies with high potential to significantly impact clinical care. The use of immediate MRI in the management of suspected scaphoid fractures was one of such technologies. In this particular intervention, cost differences per participant between groups were greater based on real-world data compared to modelling estimates. These were mainly due to behaviours from patients and clinicians alike, which were not anticipated in the economic model and ultimately increased the relative cost differences. However, regardless of the approach used, the economic evidence showed that the intervention was associated with cost savings for the NHS. This meant that basing the adoption decision solely on economic modelling evidence would have ultimately led to the same conclusion as the real-world evidence. Regardless of this individual finding, the uncertainty regarding some key model parameters supported the framework models discussed in Chapter 2, favouring the implementation of real-world studies prior to the widespread introduction of disruptive diagnostic tests.

It should also be recognised that the different methods of economic evaluation, in this case an economic model and a pragmatic RCT, can complement one another. In the case of the SMaRT trial, economic modelling informed the hypothesis testing of the subsequent pragmatic RCT and can be useful for other healthcare providers if the RCT setting is not generalisable.

6.5 Chronic headache

6.5.1 Summary of literature review

The literature review informed the modelling around the management of patients with chronic headache that were referred from primary care to either: (i) a neurologist; or (ii) direct access to brain MRI. The rationale was that the early use of MRI would reassure patients that no underlying life threatening condition, particularly brain cancer or brain aneurysms, was present. However, the prevalence of such conditions among chronic headache sufferers is almost negligible, with an estimated 0.1% prevalence of brain cancer as the underlying cause of chronic headache (Symvoulakis et al. 2007). This meant that the headache model was conceptually different from the ones built for patients with suspected scaphoid fractures and suspected colorectal cancer. In fact, the value of advanced imaging resided in the reassurance effect rather than the diagnostic information associated with the potential diagnosis of the chronic headache. This reassurance effect could in turn lead to different utilisation rates of healthcare services and ultimately to differences in the mean cost per participant in the two groups.

Howard et al. (2005) performed an RCT that showed that the use of imaging in patients with chronic daily headaches had the potential to change patient management. One important change was the reduction of referral rates to the neurology service at secondary care from 23% (17/74) in the control group to 1.3% (1/76) in the scanned group (Howard et al. 2005). Following on from this study, Thomas et al. (2010) proposed to determine whether direct access to brain imaging, in this case CT, from

primary care led to reduced referral rates to secondary care. This study was conducted in Tayside, Scotland, with an overall participation of 45% of the 309 local GPs. The findings by Thomas et al. (2010) confirmed that imaging reduced referral rates in 86% of the cases during the follow-up period (average of 1.3 years per patient). Out of the 215 patients submitted to CT, only 1.4% had a significant pathological condition, whilst 10.2% presented non-significant findings and 88.4% normal findings. Kernick and Williams (2011) assessed the impact of providing GPs with direct access to neuroimaging for patients with headache and concluded that, although the yield for clinically significant findings in neuroimaging was below 1%, the reassurance effect associated with neuroimaging remained unknown (Howard et al. 2005; Kernick and Williams 2011). Kernick and Williams (2011) added that the anxiety associated with incidental findings should not be disregarded and could, in fact, lead to increased healthcare utilisation.

Although the literature review showed that advanced imaging has the potential to decrease overall healthcare utilisation, little was known about the actual cost implications of providing GPs with direct access to brain MRI for chronic headache patients compared to conventional management with referral to neurology services. Given the limited supply of neurologists in the UK, and the potential benefits of using neuroimaging, the present subsection presents the economic model used to estimate the cost implications associated with both clinical pathways.

6.5.2 Economic modelling

A *de novo* decision tree model with a time horizon of 12 months and a healthcare payer perspective was developed to estimate the cost implications of two coexisting clinical pathways in the management of chronic headache that differed in the referral from primary care to either neurology services or direct access to brain MRI.

Figure 67 and Figure 68 illustrate, respectively, the model associated with the Neurology referral group and direct access to brain MRI group. The model structure included any primary or secondary care healthcare utilisation due to the management of chronic headache. This meant that resources that were not a consequence of the referral from primary care were included as long as they were due to chronic headache. One important example of this was attendances to the ED due to chronic headache, following which patients were considered to either be discharged without advanced imaging, underwent head CT or were admitted to hospital.

All participants in the MRI group (Figure 68) underwent an initial brain MRI scan to exclude any potential space occupying lesion, particularly brain cancer. Based on the MRI findings, three scenarios were considered:

1. Positive findings for brain cancer. Subsequently, the patient was referred to a neuro-oncology consultation and treated accordingly;

2. Clinically relevant incidental findings. Participants attended a consultation with a neurologist and subsequent follow-up – if needed – was considered. If discharged back to primary care, GPs received a full report of the MRI performed by a neuroradiologist;
3. Negative findings (no brain cancer or other significant incidental findings). The remaining patients, which included the vast majority of the cohort, followed the 'normal findings' branch. Patients were referred back to primary care with a full report of the MRI performed by a neuro-radiologist.

Table 78 and Table 79 list the branch probabilities and the unit costs included in the economic model.

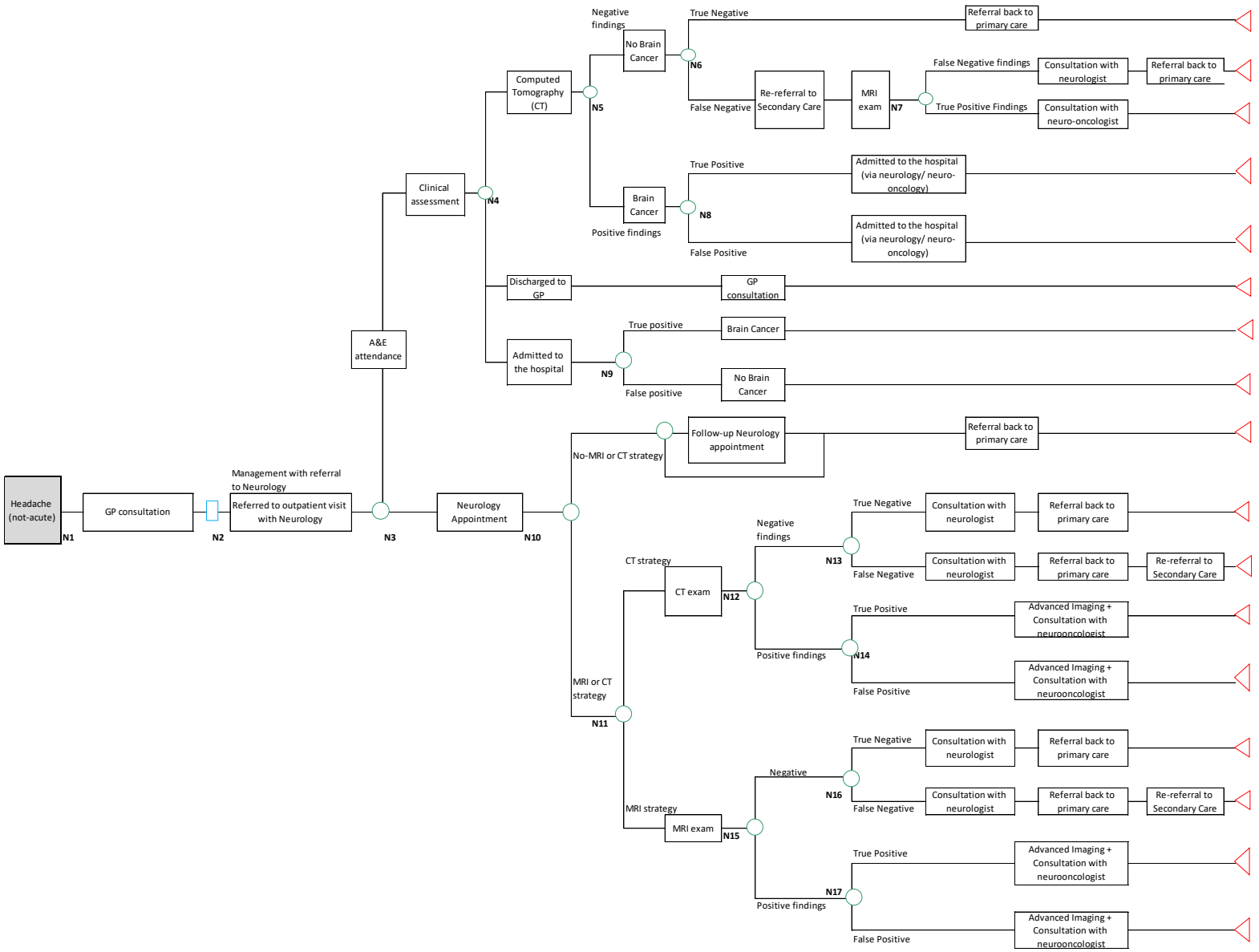


Figure 67. Economic model for patients with chronic headache managed with referral from primary care to neurology.

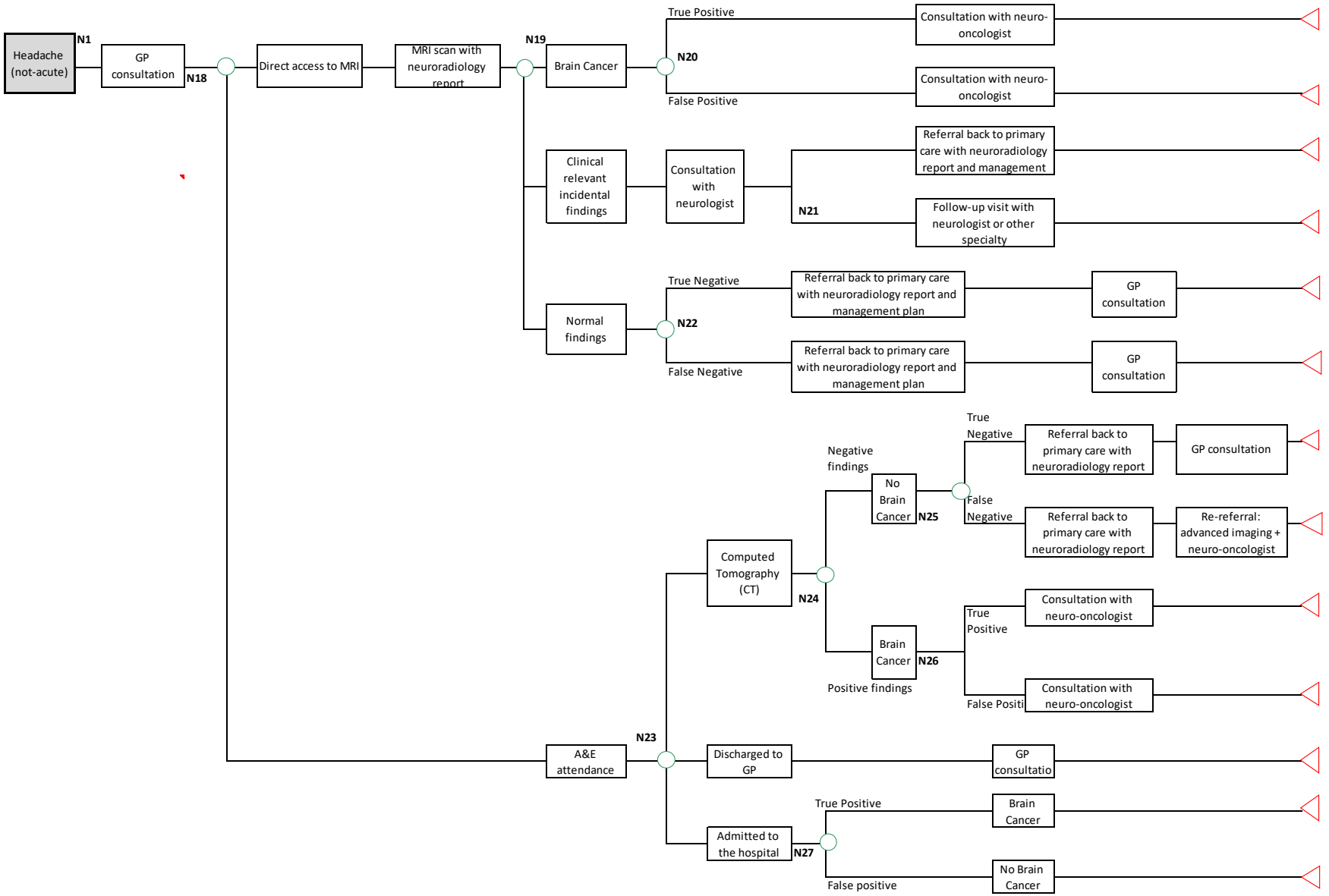


Figure 68. Economic model for patients with chronic headache managed with referral from primary care to direct access to brain MRI.

Table 78. Node probabilities included in the economic model.

Parameter	Value	Source / Description
Estimated prevalence brain cancer amongst the cohort of patients selected	0.1%	Assumption based on Symvoulakis et al. (2007)
CT sensitivity and specificity values for brain cancer	81% / 92%	Medina et al. (2010) referring to Hutter et al. (2003).
MRI sensitivity and specificity values for brain cancer	92% / 99%	
Management with referral to neurology		
Percentage of patients undergoing advanced imaging following referral to neurology	50%	Assumption based on clinical opinion from a consultant neurologist.
Percentage of patients that have a follow-up consultation with a neurologist (when no advanced imaging is used)	39.7%	Based on internal clinical audit.
Number of episodes at the ED with primary diagnosis of headache	5%	Based on internal clinical audit.
Percentage of patients attending the ED that are referred to a Computed Tomography scan (CT perfusion test)	45%	Based on internal clinical audit.
Percentage of patients admitted following presentation to the ED	5%	Based on internal clinical audit.
Percentage of patients attending the ED discharged back to primary care without advanced imaging	50%	Based on internal clinical audit.
Management with referral to brain MRI		
Percentage of incidental findings with direct access MRI	7.5%	Assumption based on clinical opinion from a consultant neuroradiologist.
Percentage of reduction of attendance to the ED amongst patients with negative MRI findings during the subsequent year	10%	It was assumed that the increase of MRI utilisation rates would halve the overall ED utilisation rate amongst patients with headache that had a MRI scan.

Table 79. Unit costs included in the economic model.

Parameter	Value	Source / Description
Primary care cost per patient with headache per 3 months	£125.20	McCrone et al. (2011). Costs inflated according to The HCHS index in Curtis (2012).
Unit cost per GP consultation	£38.50	The cost per GP consultation was £36, based on a 12 minute consultation (http://www.nice.org.uk/nicemedia/live/13076/50058/50058.pdf) and updated using Curtis (2012) data.
Primary care cost per patient with headache per 9 months with no imaging reassurance	£375.60	Estimate based on McCrone et al. (2011) and maintenance of utilisation rates and associated costs (e.g. medicines).
Percentage of reduction of primary care costs per 9 months due to a negative imaging test (either MRI or CT)	10%	Estimate based on Howard et al. (2005) and McCrone et al. (2011).
Unit cost of clinical assessment of patients with headache at the ED	£316	Estimate based on internal financial data. It was assumed the average unit cost amongst different ED episodes without admission and excluding dental care.
Unit cost per patient admitted due to headache	£1,475	National Reference Costs code PA04A (Headaches and Migraines, with CC) in National Reference Costs 2011/12.
Unit cost per head CT test (including report)	£78.70	Estimate based on internal financial data.
Unit cost per MRI head test (including report)	£148.90	Estimate based on internal financial data.
Unit cost per neurologist consultation (first visit)	£233	Estimate based on internal financial data.
Unit cost per neuro-oncologist consultation (first visit)	£233	Assumed to be equivalent to a neurologist.
Unit cost per neurologist consultation (follow-up visit)	£163	Estimate based on internal hospital costing data.
Unit cost of a patient admitted to the hospital following a positive MRI/CT for brain cancer	£2,319	National Reference Costs code AA24A (Brain Tumours or Cerebral Cysts, with CC) in National Reference Costs 2011/12.

Lifetime cost associated with brain cancer treatment	£15,086	Estimate based on the survival rates and annual costs (both for the first year, continuing year and last year of life) presented at http://costprojections.cancer.gov/annual.costs.html and assuming a linear survival rate amongst the intervals considered. A 15 year survival rate of 0% and a 3.5% discount rate were assumed to estimate the annual cost with brain cancer treatment.
Unit cost of discharge after negative findings	£0	Assumed to be included in previous procedures (e.g. in the cost of the neurologist consultation).
Unit cost per False Positive episode in MRI or CT	£381.90	It was assumed an added cost for a new MRI exam and a neuro-oncologist appointment.
Unit cost per False Negative episode in MRI or CT	£15,585	Assumed that the patient will re-enter the pathway and therefore all costs associated to its diagnosis and subsequent treatment. It was considered the addition of: (1) costs of primary care for 3 months; (2) a new MRI exam; (2) subsequent neuro-oncologist appointment; and finally (4) the annual costs of brain cancer treatment.

The results from the base case scenario are presented in Table 80. The cost per patient in the Neurology and MRI group were estimated at £557 and £542, respectively. The economic model showed that direct access to brain MRI was likely to produce marginal cost savings of £15.6 per participant, equivalent to a reduction of 3% in the total 12-month healthcare costs. In contrast, if only secondary care costs were considered, the MRI pathway was likely to generate a higher proportion of cost savings to the healthcare payer, estimated at 43% of the absolute unit cost per patient in the neurology pathway. Most healthcare costs were estimated to reside in primary care, particularly in the MRI group as only participants with clinically relevant findings in the MRI were assumed to need further follow-up at secondary care.

6.5.3 Comparative analysis: real-world versus model results

Results from the chronic headache observational study were presented in Chapter 4. The use of direct access to MRI led to cost savings at both 6 and 12-month post-recruitment (£333 and £518, respectively). Hence, real-world evidence showed that the actual cost difference per participant was much higher than anticipated in the economic model (Table 80).

.

Table 80. Mean cost per participant from the economic model and observational study for both the Neurology and the MRI group.

	Economic model data (12 months)		**Real-world data (observational study)**			
	Cost per patient	Potential savings	6 months		12 months	
			Cost per patient	Potential savings	Cost per patient	Potential savings
Neurology Group	**£557**	**£15**	**£578**	**£333**	**£808**	**£519**
Direct access to brain MRI	**£542**		**£245**		**£289**	

The comparison of model-based and real-world evidence showed that modelling costs in the Neurology group were underestimated whilst costs in the MRI group were overestimated. These cost differences were due to multiple reasons, both in terms of the valuation of key parameters as well as the overall model structure.

In the Neurology group, parameter inconsistencies were identified between the model data and real-world clinical practice. These inconsistencies related to parameter estimates rather than the actual structure of the model. First, primary care utilisation was overestimated in the economic model. In practice, referral to neurology services led to a substantial decrease in GP appointments and consequent costs (decrease of mean number of 1.5 appointments in the 12 months post-recruitment vs 12 months pre-recruitment). This was the only model parameter that contributed to an overestimate of total costs in the Neurology group.

Second, treatment options for chronic headache were not considered. This meant that medication costs and procedures such as nerve block injection and Botox treatment were not included, thus leading to an important underestimate of costs. The observational data found that patients in the Neurology group underwent a mean of 0.30 Botox or nerve injection procedures. Given the high-cost of each treatment session (unit cost per nerve block injection session of £587-£636 or Botox treatment of £650-£749) (NHS Improvement 2018), and the difference in utilisation between the two groups (mean 0.30 vs 0.05 in the Neurology and MRI group, respectively), this led to an important underestimation of costs.

Third, the utilisation of advanced imaging in the Neurology group, estimated at 50%, was 10% higher (78 scans in 129 participants). Furthermore, out of these 78 scans, there was only one CT, with the remaining 77 scans being head MRI, a more expensive scan. Fourth, the economic model assumed that out of the patients referred to Neurology where no advanced imaging is considered (branch 'No MRI or CT strategy'), only 39.7% had at least one follow-up appointment. This estimate was not corroborated by real-world data as 77% (41 out of 53) participants had at least one follow-up appointment. Furthermore, out of these participants, 58% (24 out of 41) participants had more than

one follow-up appointment, a situation that was not contemplated in the economic model. These elements combined, led in the Neurology group, to a mean cost per patient underestimate of £251 (£557 vs £808) in the economic model compared to the real-world data.

In the MRI group, both parameter and structural differences were identified in the comparison of model data with empirical evidence. The main parameter difference derived from the utilisation of primary care resources. Participants in the MRI group were assumed to have a much higher consumption of GP face-to-face appointments compared to observed data. The reason was that these patients would have no formal follow-up at secondary care so, following the head MRI, the ongoing management of chronic headache was assumed to take place in primary care. However, following recruitment to the study, a mean number of 1.19 GP appointments per MRI participant was observed (vs 1.82 in the Neurology group). This meant that, despite not having formal follow-up at secondary care, some MRI participants did not even attend primary care, having the MRI results communicated by GPs either over the telephone (lower unit cost compared to face-to-face appointment) or via post. The reduction in the overall utilisation of primary and secondary care resources following a negative MRI scan confirmed the reassurance effect proposed in the clinical and economic literature. This reassurance effect also translated into lower than anticipated utilisation rates of both inpatient and emergency care, thus contributing to the overestimation of costs in the MRI group in the economic model (£542 vs £289).

The structure of the economic model for participants in the MRI group assumed those with normal MRI findings ('True Negative' findings in Node 22 in Figure 68) were discharged back to primary care where the management of chronic headache would occur. This did not accurately reflect routine clinical practice as 18% (17/95) participants were re-referred to secondary care, this time for an appointment with a neurologist. Figure 69 depicts the changes made to the structure of the economic model. All 88 participants in the MRI group with true negative findings (7 participants had incidental findings) were referred back to primary care. Out of these, 44% (39/88) of participants had no GP appointment and the remaining 56% (49/88) had at least one GP appointment. From participants with at least one GP appointment, 35% (17/49) were subsequently re-referred to secondary care, this time for a neurologist appointment. These were referrals associated with patients that, despite the negative MRI, presented higher headache burden compared to the rest of MRI participants. All branch probabilities described above are also illustrated in Figure 69.

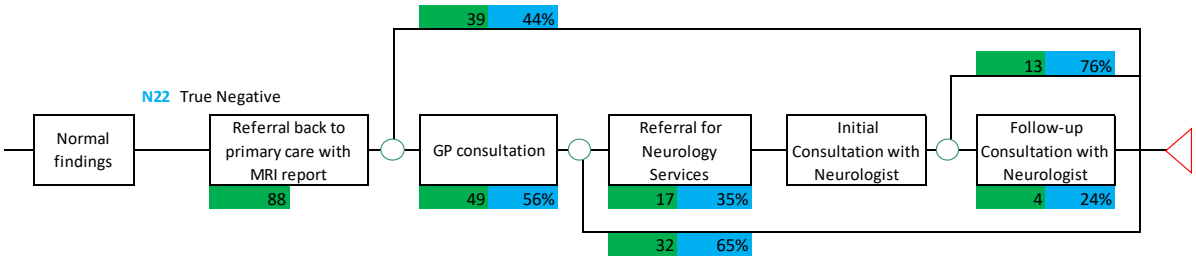


Figure 69. Changes to the economic model based on data from the observational study (blue cells indicate the respective branch probability and absolute number of patients based on real-world data).

6.5.4 Potential implications

Although the two pathways being evaluated had recently been constituted as standard care for the management of chronic headache patients, there was little evidence concerning the cost implications of each pathway. Hence, given the unknown cost implications, and also the increasing volume of chronic headache patients (population at risk) being referred from primary care, it was decided to conduct an observational, prospective study to support future decision making processes.

Data from this observational study and the *a priori* economic modelling were then compared and important differences were identified. In summary, the *a priori* model underestimated the costs associated with provision of care for participants in the Neurology group and overestimated the costs in the MRI group. This combination meant that the economic model suggested that direct access to MRI compared to management with referral to neurology services would only produce marginal cost savings (£16 cost difference per participant) as opposed to the significant cost savings identified in the observational study (adjusted 12-month cost differences between groups of over £500 per participant, $p<0.001$). The latter meant that basing the adoption decision of direct access to MRI exclusively on modelling data might have misled decision makers and potentially negatively affect the uptake of direct access to brain MRI in the context of managing chronic headache patients.

6.6 Suspected colorectal cancer

6.6.1 Summary of literature review

The literature review informed the parameter values and the model structure of the utilisation of Computed Tomography Colonography (CTC) or Optical Colonoscopy (OC) as the initial investigation of suspected colorectal cancer (CRC). As with the other models presented in this chapter, the literature review synthesised the prevalence of the clinical condition (medium to large polyps and CRC) and the accuracy of both imaging tests in the diagnosis of the clinical condition.

The prevalence rate of CRC was estimated at 3.7% based on historical internal hospital data which was also consistent with the findings of a systematic review and meta-analysis (Pickhardt et al. 2011). A 10% prevalence of large polyps was estimated based on a systematic review and meta-analysis (Halligan et al. 2005).

The accuracy of both imaging tests, CTC and OC, in the diagnosis of CRC and large polyps was estimated. OC had a 94.7% sensitivity value for the detection of CRC (95% CI: 90.4%, 97.2%) (Pickhardt et al. 2011) and a 94% sensitivity for large polyps (i.e. diameter over 10 mm) (Menardo 2004).

CTC has high accuracy levels in the detection of large and medium polyps and is especially sensitive to the detection of symptomatic CRC (Halligan et al. 2005). Systematic reviews showed that sensitivity and specificity of CTC increases with larger polyps (NICE 2014). According to a meta-analysis by Halligan et al. (2005), CTC had, respectively, 93% (95% CI: 73%, 98%) and 97% (95% CI: 95%, 99%) sensitivity and specificity values for large polyps. In the diagnosis of CRC, CTC was shown to present a 95.9% sensitivity value (95% CI: 91.4%, 98.5%) (Halligan et al. 2005). The latter value was corroborated by Pickhardt et al. (2011), who estimated a sensitivity value of 96.1% (95% CI: 93.8%, 97.7%). The base case scenario assumed the above mentioned values. Older studies tend to underestimate the accuracy levels of CTC as new generation CT scanners are likely to significantly improve overall accuracy levels in the diagnosis of both large polyps and CRC.

NICE (2014) acknowledged that OC holds the highest clinical accuracy for diagnosis of CRC and therefore should be considered the gold-standard as a first line examination test. However, this may no longer be the case as a recent study showed that no CRC was missed by OC and only one with CTC (1 out of 29 patients with CRC, i.e. an average sensitivity of 96.5%) (Atkin et al. 2013). The authors concluded that the widespread use of CTC as an alternative to OC was justified, provided that suitable guidelines and training schemes were put in place (Atkin et al. 2013).

In summary, CTC is highly sensitive, particularly for CRC. Given the relatively low prevalence of CRC, CTC could be considered as a suitable first line investigation test for low to medium risk symptomatic patients, as a direct alternative to OC.

A critical appraisal conducted by the Centre for Reviews and Dissemination at the University of York concluded that findings from cost-effectiveness studies varied greatly in terms of comparative analysis of costs and effects associated with CTC and OC (National Institute for Health Research 2009). In addition, most economic evaluations were related to screening programmes and not, as considered in this evaluation, to symptomatic patients. Nevertheless, this centre highlighted that the relative cost of CTC and OC were two critical parameters to the incremental cost-effectiveness ratio. Their findings suggested that CTC is likely to be cost-effective if it presents a unit cost between 22% and 52% compared to OC (National Institute for Health Research, 2009).

6.6.2 Economic modelling

A *de novo* decision tree model with a time horizon of 6 months and a healthcare payer perspective was developed to estimate the cost implications of CTC as a direct alternative to OC in patients with low to intermediate risk of CRC. Figure 70 and Figure 71 illustrate the model structure associated with the OC and CTC groups, respectively. The incidence and accuracy parameters are summarised below in Table 81.

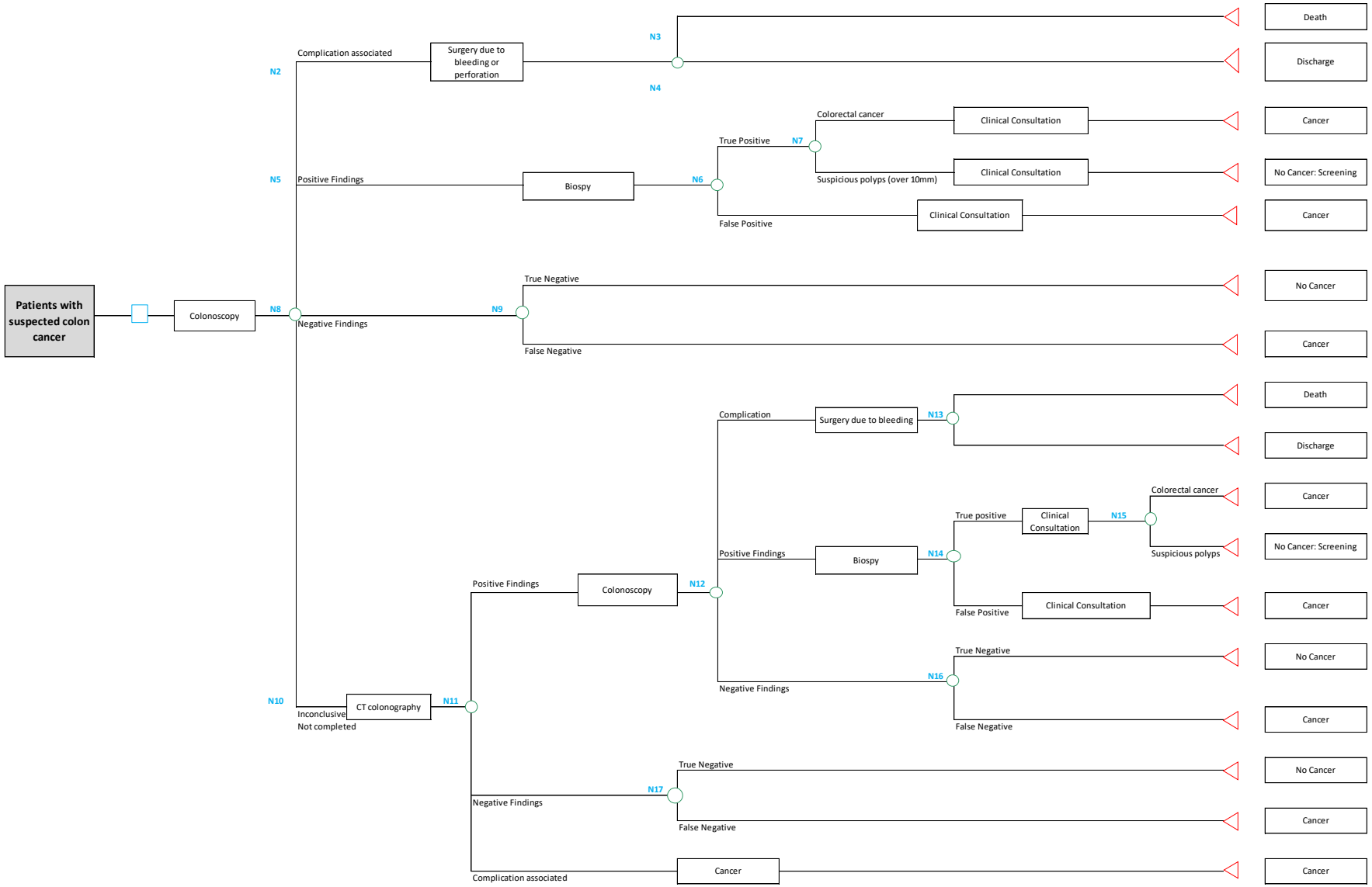


Figure 70. Short-term model for participants in the OC group.

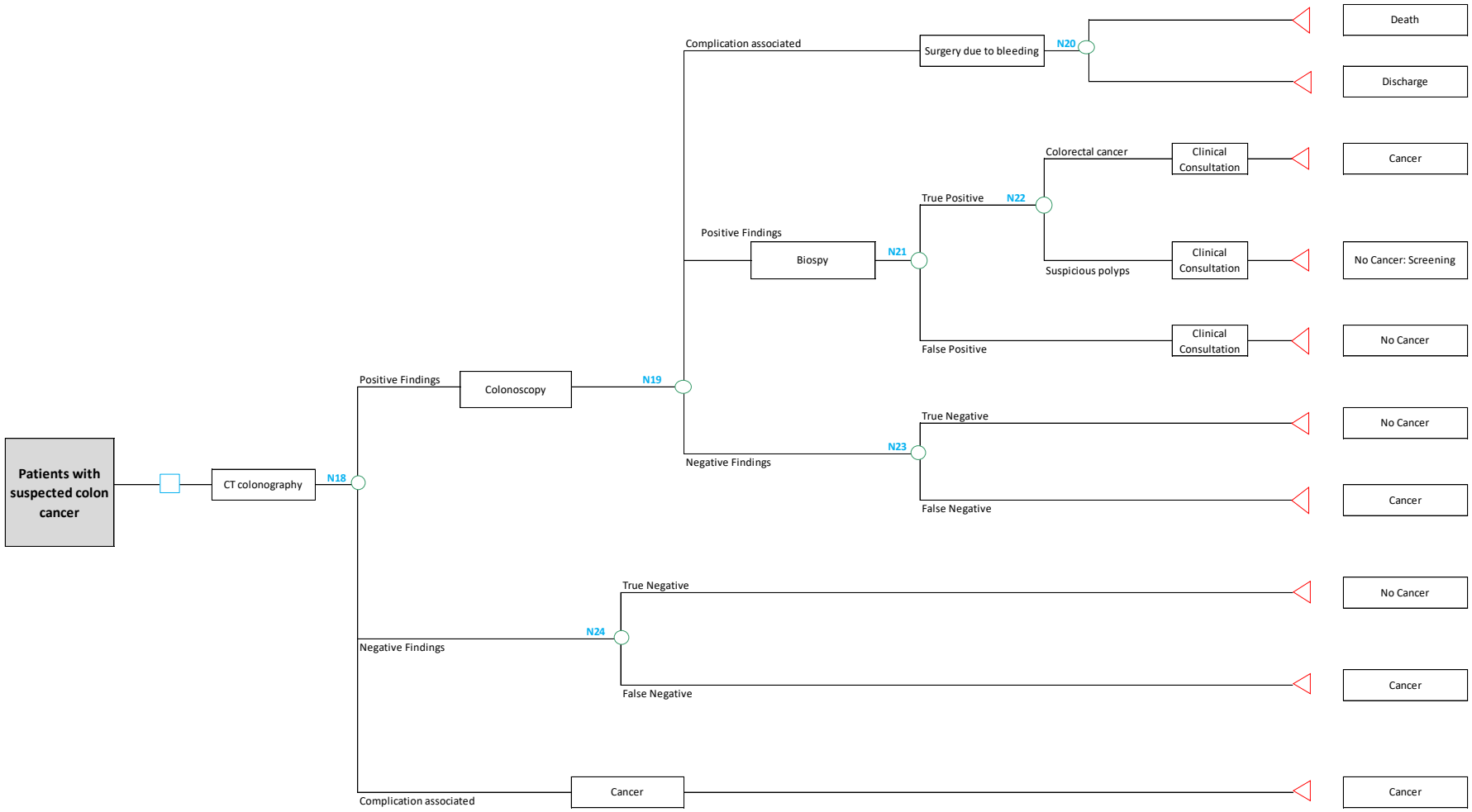


Figure 71. Short-term model for participants in the CTC group.

Table 81. Incidence, sensitivity and specificity parameters included in the economic model.

Parameter	Value	Source / Description
Estimated prevalence of CRC	3.7%	Based on internal hospital data and clinical studies.
OC sensitivity and specificity to detect CRC	Sens: 94.7% Spec: 95.0%	Based on Pickhardt et al. (2011).
OC sensitivity and specificity to detect large polyps	Sens: 94.0% Spec: 95.0%	Sensitivity based on Menardo (2004). Specificity assumed to be equal to CRC.
CTC sensitivity and specificity to detect CRC	Sens: 96.1% Spec: 95.0%	Based on Pickhardt et al. (2011).
CTC sensitivity and specificity to detect large polyps	Sens: 93.0% Spec: 97.0%	Based on Halligan et al. (2005).

Table 82 summarises the different node probabilities and unit costs included in the economic model. All data presented were obtained from either: literature review; national reference costs; clinical experts; or internal hospital data.

Table 82. Node probabilities and unit costs included in the economic model.

Node (Fig. 6 , Fig. 7)	Parameter	Value	Source / Description
N2/N12	Probability of complication (bleeding due to perforation) associated with OC	0.25%	Information retrieved from literature (Rabeneck et al. 2008).
N3/N13	Probability of death associated with a surgery due to bleeding of perforation	3.0%	
N10	Inconclusive colonoscopy / patients unable to comply	7.5%	Retrieved from an internal clinical audit with 862 patients.
N10	Inconclusive colonoscopy after a positive CTC.	0%	Assumption. The rationale was that patients would find bowel preparation easier to follow and more tolerable, thus leading to a lower proportion of inconclusive colonoscopies.
N18	Added probability of cancer incidence due to CTC scan.	0.04%	Gonzalez et al. (2009).

N/A	Unit cost of OC	£527.20	NHS Reference Costs, code FZ51Z - Diagnostic Colonoscopy, 19 years and over.
N/A	Unit cost of CTC	£242.20	NHS Reference Costs, unbundled CTC.
N/A	Unit cost of surgery due to bleeding	£6,819	NHS Reference Costs, code FZ76A - Distal Colon Procedures, 19 years and over with Major CC.
N/A	Unit cost of death	£0	Assumption. No cost was considered in the base case scenario. No sensitivity analysis was performed as this event probability was negligible (expert opinion).
N/A	Unit cost of discharge without further investigation.	£0	Assumption. Cost assumed to be included within the surgery episode cost.
N/A	Cost of biopsy	£43.20	Assumption. Cost difference between the unit cost for a colonoscopy with and without biopsy (using National Reference Codes FZ52Z- Diagnostic Colonoscopy with Biopsy, 19 years and over and code FZ51Z - Diagnostic Colonoscopy, 19 years and over.
N/A	Initial outpatient appointment	£132	Internal financial data. Any patient with a positive finding (either true positive or false positive will have a follow-up consultation).
N/A	Lifetime cost of cancer due to CTC	£25,500	Cost of Cancer estimated from NHS England (2011).
N/A	Added cost in true positive patients	£0	No added cost considered.
N/A	Added cost in false positive patients	£527.20	It was considered the added cost of performing a serial colonoscopy without biopsy.
N/A	Added cost in true negative patients	£0	No added cost considered.
N/A	Added cost in false negative patients	£570.50	Added cost of a repeated colonoscopy with biopsy.

Table 83 presents the unit cost per participant for each strategy considered. The mean cost per participant in the OC and CTC were, respectively, £660 and £393. The use of CTC, instead of OC, as the initial colonic investigation led to potential savings per participant of £267 or 40% compared to the mean cost per participant in the OC group.

Table 83. Mean cost per participant from the base case scenario for the OC and CTC groups.

Strategy		Cost per patient	Potential savings
OC	OC as the initial imaging modality	**£660**	**£267**
CTC	CTC as the initial imaging modality	**£393**	

Sensitivity analyses:

Deterministic sensitivity analyses were performed on five model parameters, based on a minimum and maximum range (Table 84). First, the prevalence of CRC was varied to a minimum of 1% and a maximum of 10% (baseline 3.7%). Second, sensitivity and specificity for CTC were decreased to 80% and increased to 100% (i.e. no false negative of false positive findings). Third, the incidence of CTC cancer-induced was increased to 0.1% and 1.0% from a baseline value of 0.035%. Fourth and fifth, the unit cost of CTC and OC was varied ±50%.

Table 84. Minimum and maximum values considered in the five deterministic sensitivity analyses.

Parameter	Base case value	Range of values (min-max)
1. Prevalence of CRC amongst the initial cohort of symptomatic patients	3.7%	1%-10%
2. Sensitivity and specificity of CTC for the diagnosis of CRC	96.1% / 95.0%	80%/80% - 100%/100%
3. Increased cancer incidence due to CTC radiation	0.035%	0.1-1%
4. Unit cost per CTC	£242.20	±50% (£121.1; £363,3)
5. Unit cost per Optical Colonoscopy	£527.40	±50% (£263.6; £790.9)

Table 85 summarises the results from the base case scenario and all five sensitivity analyses. In all analyses, the use of CTC as the initial diagnostic colonic test compared to OC was associated with cost savings to the healthcare payer (negative values represent cost savings). First, the increase in the prevalence of CRC led to a decrease in cost savings. This was due to the fact that participants in

the CTC group that have positive findings will undergo a subsequent OC (second test), thus increasing the overall costs for this group. Second and third, any improvement on the diagnostic test's accuracy, either CTC or OC, will lead to an improvement in the respective costs of such group. Fourth and fifth, any changes in the unit cost of CTC or OC will lead to changes in the mean cost difference. For instance, any increase to the unit cost of CTC led to a decrease in the cost savings. However, even if the unit cost of CTC increased by 50% (to a unit cost of £363), using CTC as the primary investigative scan still led to cost savings per participant of £156.

Table 85. Difference in mean cost per participant based on five deterministic sensitivity analyses.

Parameter	Minimum	Maximum
Base case scenario	-£267	
1. Prevalence of CRC amongst the initial cohort of symptomatic patients	-£280	-£241
2. Sensitivity and specificity of CTC for the diagnosis of CRC	-£192	-£293
3. Increased cancer incidence due to CTC radiation	-£254	-£58
4. Unit cost per CTC	-£380	-£156
5. Unit cost per Optical Colonoscopy	-£33	-£503

6.6.3 Comparative analysis: real-world versus model results

Results from the observational clinical study were presented in Chapter 5. The intervention with CTC as the primary colonic investigation led to cost savings per participant of £345, compared to the £267 figure estimated from the economic model (Table 86). Hence, real-world evidence showed that the actual cost difference per participants was higher than anticipated via the economic model.

Table 86. Mean cost per participant from the economic model and observational study for both groups.

Strategy		Economic Model		Observational study	
		Cost per patient	Potential savings	Cost per patient	Potential savings
OC	OC as the initial imaging modality	**£660**	**£267**	**£991**	**£345**
CTC	CTC as the initial imaging modality	**£393**		**£646**	

In both groups, the economic model underestimated the costs associated with the provision of healthcare compared to real-world data although the cost differences were relatively close. These cost differences were due to multiple reasons, both in terms of the valuation of key parameters as well as the overall model structure, both for the OC and the CTC groups.

In the OC group, the structure of the model accurately reflected real-world clinical practice as per the observational study. However, two parameters contributed to comparatively lower costs of £331 (£991 - £660) per participant. First, the probability of the initial OC test being inconclusive or patients being unable to comply, estimated at 7.5% in the economic model was found to be marginally higher (11.4%). This meant that a higher proportion of participants underwent a second test, thus increasing costs. Second, and more importantly, the unit cost of OC was found to be significantly higher than the model assumption (£527) and variable according to different purposes (e.g. analgesia, upper and lower gastrointestinal test done simultaneously, ranging from £515 to £911). Furthermore, as with the CTC group, part of the increase in costs was explained by the four year gap between the economic model data (2013-14) and the real-world evidence (2017-18). Costs from the economic model were updated using the HCHS index (Curtis and Burns 2018), leading to a cost difference between groups in the economic model (from £267 to £283) that was closer to the estimates from the real-world trial (£345).

In the CTC group, the economic model represented an oversimplification of real-world clinical practice that led to an underestimate of the CTC unit cost per participant of £252 (£646 - £393). This was due to two reasons. First, as with the OC group, the CTC scan can produce inconclusive findings or patients may be unable to comply (e.g. unexpected claustrophobia). In both cases, this would lead to a second diagnostic test, a situation that was not considered in the economic model despite representing 8.8% of all CTC scans. Second, and unlike OC, CTC visualises extra colonic organs. This led to the identification of extra colonic findings in 25% of CTC scans, with 4.4% of CTC scans leading to changes in the care management pathway. This again represented an oversimplification of the model that produced to lower cost estimates. Figure 72 illustrates the two branches added to the CTC model to correct these two limitations.

In both groups, real-world evidence did not find evidence of complications (e.g. bleeding due to bowel perforation, adverse reactions to bowel preparation) associated with the OC or CTC scan. However, although rare, these events do occur and therefore the model structure appropriately reflects clinical practice.

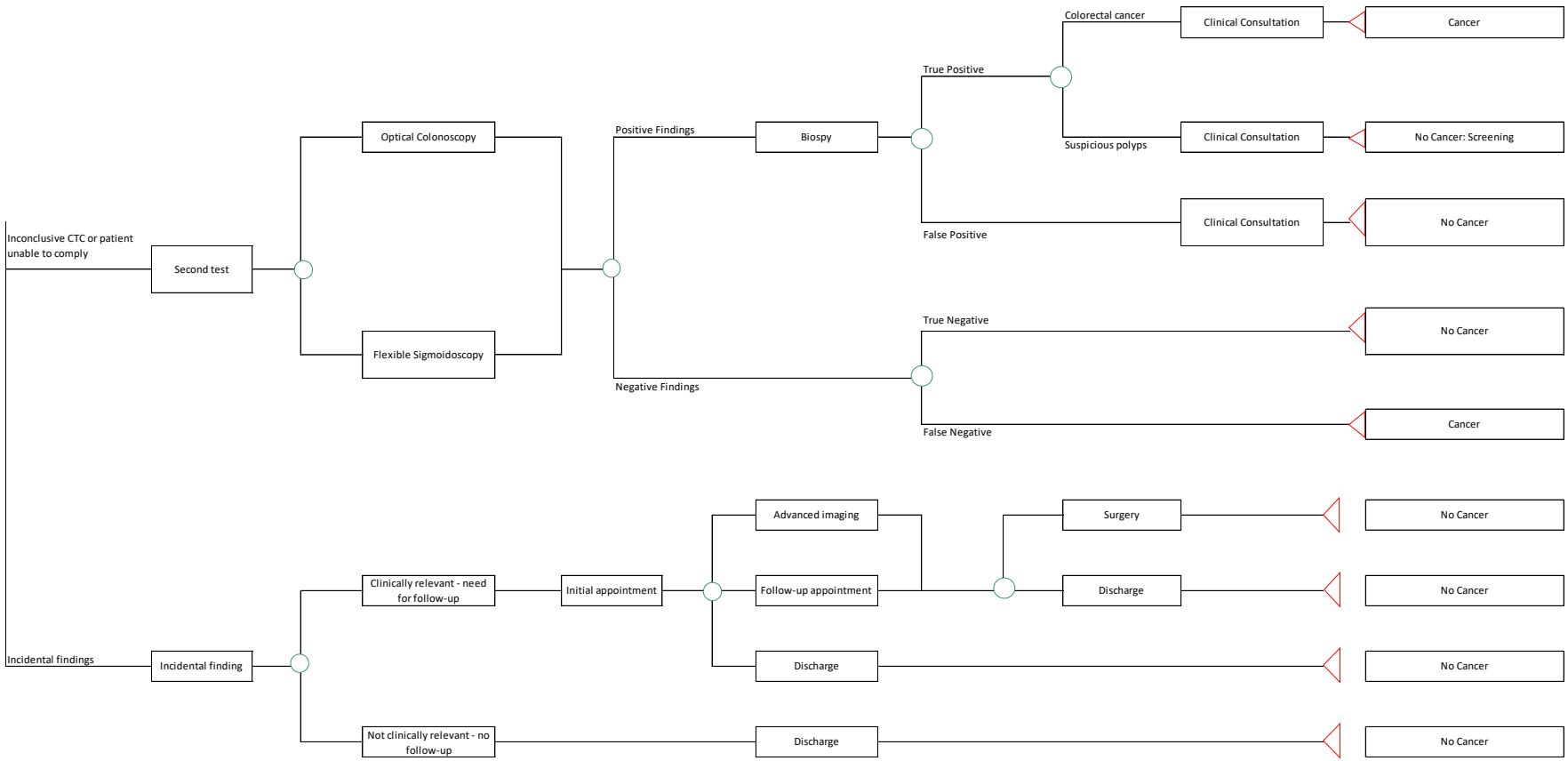


Figure 72. Changes to the structure of the economic model in the CTC group based on the empirical evidence from the colon study (two branches added to the original model).

6.7 Discussion

The potential implications of using model-based data compared to real-world evidence were explored across three clinical pathways. The literature review showed that the type of data informing the technology's adoption decision depends on the technology itself, particularly its value proposition and likelihood to impact routine clinical practice, and the level of existing clinical and economic evidence. The more disruptive the technology, the less likely it is to have published clinical and economic evidence, thus highlighting the need for real-world evidence.

This was the case for using immediate MRI in the management of suspected scaphoid fractures. The comparative analysis of data from the *a priori* economic model and the RCT showed that both methods supported the adoption of the intervention. However, the economic model represented an oversimplication of real-world clinical practice that could have led to differences in the adoption process.

In contrast, if the technology's anticipated benefits, both in terms of costs and effect, are documented in previous real-world evidence, economic modelling can provide a solid framework for evidence synthesis on which to base a funding decision. This was the case in the management of suspected colorectal cancer, where economic modelling, despite underestimating the costs in both groups, led to similar cost differences between groups and would have led to the same conclusion for decision makers. The latter was due to the existence of real-world data around the key model parameters, the probability for subsequent colonic investigations following the initial test (either CTC or OC) and the unit costs of both tests, particularly CTC.

A different scenario occurred in the evaluation of direct access to MRI for patients with chronic headache. In that case, the *a priori* economic model failed to represent real-world clinical practice, leading to substantial cost differences between groups. Ultimately, basing the decision solely on the economic model could have led to the non-adoption of direct access to MRI. However, given the potential clinical and cost implications, as well as the population at risk (over 500 patients/year), it was decided to conduct a prospective, observational study which ultimately showed the existence of statistically significant cost savings associated with direct access to MRI.

As a corollary, the combination of model-based and trial-based economic evaluations should be considered to assist decision makers in the context of different interventions and clinical settings.

Chapter 7. General discussion

7.1 Chapter overview

This chapter discusses the aims and key findings of each study included in the thesis. Subsequently, a detailed description is given regarding the implementation work completed to redesign each of the clinical pathways as well as its real-world impact at GSTT and broader implications for further research and healthcare policy. The chapter ends with the strengths and limitations of the PhD and the overall conclusion of the thesis.

7.2 Aims of the thesis

Chapter 2 presented the background economics on the new or novel utilisation of advanced imaging. In summary, the combination of demand in growth for advanced imaging with technological developments that led to the simultaneous increase in diagnostic accuracy and decrease in scanning acquisition and processing times, meant that advanced imaging became more accurate, accessible and cheaper. However, there is little evidence on the cost and health economic implications associated with the use of advanced imaging.

The overarching aim of the PhD was to evaluate the utilisation of advanced imaging across multiple clinical pathways. In the context of NHS financial constraints, the evidence from this PhD challenges the established diagnostic paradigm that advanced imaging, being more expensive and less available than basic imaging modalities (e.g. radiographs, ultrasound), should be reserved until later stages of diagnostic pathways. Although seemingly counter intuitive, it was hypothesised that the upfront utilisation of an expensive but more accurate diagnostic test would streamline the subsequent diagnostic and, if needed, treatment pathway, leading to an overall decrease in total healthcare costs.

For this purpose, one prospective randomised clinical trial and two prospective observational studies were conducted to investigate the use of advanced imaging in patients with suspected scaphoid fracture, chronic headache and suspected colorectal cancer. For all three interventions, the condition-specific costs from the NHS perspective were considered as the primary outcome. The rationale for this approach was three-fold. First, the use of advanced imaging was hypothesised to optimise the use of healthcare resources thereby promoting efficiency and avoiding waste, thus contributing to the NHS sustainability agenda. Second, the interventions were hypothesised to also improve clinical outcomes and self-reported patient satisfaction levels. Third, despite the improvement in both clinical outcomes and costs, it was not deemed feasible to power the study based on other outcomes (e.g. difference in scaphoid non-union fractures). This was not considered feasible within the scope of the TOHETI programme and this PhD.

The clinical studies evaluated the cost and clinical implications of the early use of advanced imaging in three specific clinical pathways for patients with: (i) suspected scaphoid pathway (chapter 3); (ii) chronic headache (chapter 4); and (iii) suspected colorectal cancer (chapter 5). Albeit different in nature, as the first two interventions examined the use of MRI and the third CT, the rationale was consistent, i.e. to impact clinical pathways by promoting optimal care based on the innovative use of advanced imaging. The following subsections summarise the research findings from the economic models produced in the initial stages of this PhD and the subsequent real-world studies conducted (detailed in Chapter 6) and particular attention is given to the implementation work completed to develop a new clinical pathway (suspected scaphoid fracture pathway) and promote the uptake of existing clinical pathways (chronic headache and suspected colorectal cancer pathways).

7.3 Research and implementation work

This subsection is organised as per the remaining chapters of the thesis, addressing the three clinical pathways investigated. This subsection describes first the research conducted (modelling and empirical studies) and, second, the implementation of changes to clinical practice for each pathway. Third, for all three clinical pathways, the real-world implications for clinical practice at GSTT are quantitatively estimated.

7.3.1 Suspected scaphoid fracture

Chapter 3 presented the clinical and economic evidence around the use of immediate MRI in the management of suspected scaphoid fractures compared to standard care, which relies on the use of radiographs only during the acute episode.

Model-based evaluation:

A systematic literature reviewed the evidence around the use of immediate MRI in the management of suspected scaphoid fractures. Given the intervention's innovative nature, there was no empirical data and economic modelling data presented important limitations and showed conflicting evidence. Furthermore, most economic models considered the utilisation of MRI a few days after the acute presentation, rather than at presentation, and also did not consider subsequent changes in care management based on the MRI findings. Based on this systematic literature review, it was considered that no evidence reflected the evaluation of using immediate MRI in the management of suspected scaphoid fractures and therefore a *de novo* decision tree model was developed.

A 3-month time horizon model compared the total healthcare costs of the intervention with immediate MRI for patients with negative findings in the initial scaphoid radiographs with standard care, which relied on the use of radiographs only. Furthermore, subsequent care in the intervention group was streamlined based on the MRI findings. Participants with positive findings for fractures were directly referred to the fracture clinic where further follow-up and treatment was considered: a proportion of

patients with negative findings were discharged without any follow-up; and the remaining were considered to come back to hospital for further assessment. The baseline scenario estimated that the intervention with immediate MRI would lead to a 24% 3-month cost reduction compared to standard care, equivalent to £109 per participant or over £74,000 annually based on the GSTT historical activity data. Furthermore, the deterministic sensitivity analyses showed that even when the incidence of scaphoid fractures was doubled or even if all patients with negative MRI findings were followed-up at fracture clinic, the intervention still remained cost saving.

Despite this favourable economic evidence, due to the absence of observed data and the uncertainty surrounding the real-world implementation of such an innovative intervention, we established the need for empirical data. This decision was consistent with existing frameworks for the evaluation of medical imaging technologies, e.g. Gazelle et al. (2011) (see Chapter 2 for further detail), and was based on three elements. First, the at-risk population, with over 1,000 yearly presentations to GSTT's ED due to suspected scaphoid fractures. Second, the clinical impact was considered to be large as the intervention fundamentally changes the diagnostic and treatment pathway. Third, although the potential economic impact for the NHS was not large, the intervention was believed to lead to cost savings to the healthcare payer. For these three reasons, a single-centre pragmatic randomised trial was conducted at GSTT.

Trial-based evaluation (SMaRT trial):

Based on guidance from evaluation frameworks for imaging tests and the absence of robust clinical and economic evidence around the intervention, we deemed the use of high-level evidence to be appropriate. Consequently, we undertook the SMaRT trial, a pragmatic, randomised clinical trial that recruited 136 participants and analysed the total NHS resource utilisation over a period of 6 months. The latest evaluation frameworks supported the assessment of the intervention up to a societal level (Gazelle et al. 2011; Fryback and Thornbury 1991). Hence, the SMaRT trial also included the assessment of costs from a societal perspective as a secondary analysis. However, consistent with the recommendations by NICE (Crabb 2011), the present study adopted the NHS perspective for costs in order to provide evidence relevant to the UK clinical practice.

Primary research hypothesis. What are the economic and clinical benefits of using immediate wrist MRI in the acute management of suspected scaphoid fractures?

The primary hypothesis was not rejected as there was no statistically significant cost difference at 3 months post-recruitment (Table 87). However, when this cost analysis was extended up to 6 months post-recruitment (first secondary hypothesis), the use of immediate MRI showed statistically significant cost savings between groups. A range of operational and clinical outcomes were also included as secondary hypotheses. Some clinical outcomes were intermediate outcomes, such as diagnostic

accuracy, whilst others, like self-reported quality of life and patient satisfaction, were final outcomes. This allowed simultaneous understanding of the operational link between the diagnosis and the treatment decision whilst addressing uncertainty between the intervention and tangible final clinical outcomes (see chapter 2 for further methodological details).

Table 87. Primary hypothesis considered in the SMaRT trial.

Primary Null Hypothesis	Null hypothesis rejected / not rejected
There is no difference in 3-month NHS cost per patient between the use of immediate MRI in the management of suspected scaphoid fractures with negative findings in the initial radiograph compared to standard care.	Null hypothesis not rejected. There was no significant difference in total costs between both groups (-£174, 95% CI: - £378 to £30, p=0.094).

The results from the SMaRT trial indicated that the immediate MRI was associated with cost savings at 6 months post-recruitment and was highly likely to be cost-effective at month 3 and 6 post-recruitment given conventional willingness-to-pay thresholds. With regards to non-economic outcomes, the intervention produced quicker and more accurate diagnosis, leading to an improved and streamlined treatment pathway. Contrary to previous evidence from a systematic literature review, the use of advanced imaging did not produce significant difference in the time off work or hand immobilisation with plaster cast. The latter seems to be due to the fact that the use of MRI led to a higher number of fractures being detected ($p<0.05$).

The findings from the SMaRT trial led to the adoption of immediate wrist MRI as part of standard care at GSTT. The implementation work conducted as part of the TOHETI programme to incorporate the intervention as part of routine care is presented below.

Implementation work:

Main challenges:

The translation of research findings into routine clinical practice is challenging, with some evidence suggesting that it takes too long and quite often it does not happen at all. Balas and Boren (2000) estimated that scientific discoveries took around 17 years to be integrated into clinical and community practice. Not only does it take too long, they also estimated that only about 14% of research discoveries are translated into benefits to patient care (Balas and Boren 2000). Although these figures are skewed by the research pipeline associated with innovative interventions, in particular pharmaceutical drugs, translational research still remains a challenging field. In fact, despite the different methodological approaches suggested in the translational research field, from two (T1-T2) to five stages (T0-T4)

frameworks (see 2.3.3), there is wide acceptance around the overarching concept of translational research and its increasing importance (Zoellner and Porter, 2017).

Conscious of these challenges, a two-fold strategy to implement the intervention as part of routine clinical practice was considered. First, the intervention, i.e. the immediate use of MRI in an acute setting, represents a paradigm shift as MRI is not used in the acute setting. For this reason, a significant element of TOHETI was to transform the work across Clinical Imaging and Medical Physics (CLIMP) directorate, i.e. the Radiology Department. In fact, it was deemed essential to change the 'ways we work' in order to implement and sustain changes as part of the day-to-day job of all professionals, from clinical staff (e.g. radiologists, radiographers) to administrative and managerial roles. Second, although focused in the acute setting, the intervention fundamentally changed the entire diagnostic and treatment pathway. As with most clinical pathways, patients with suspected scaphoid fractures crossed not only the imaging department but multiple clinical directorates (Figure 73). This second level of dissemination thus included multiple stakeholders from different directorates.

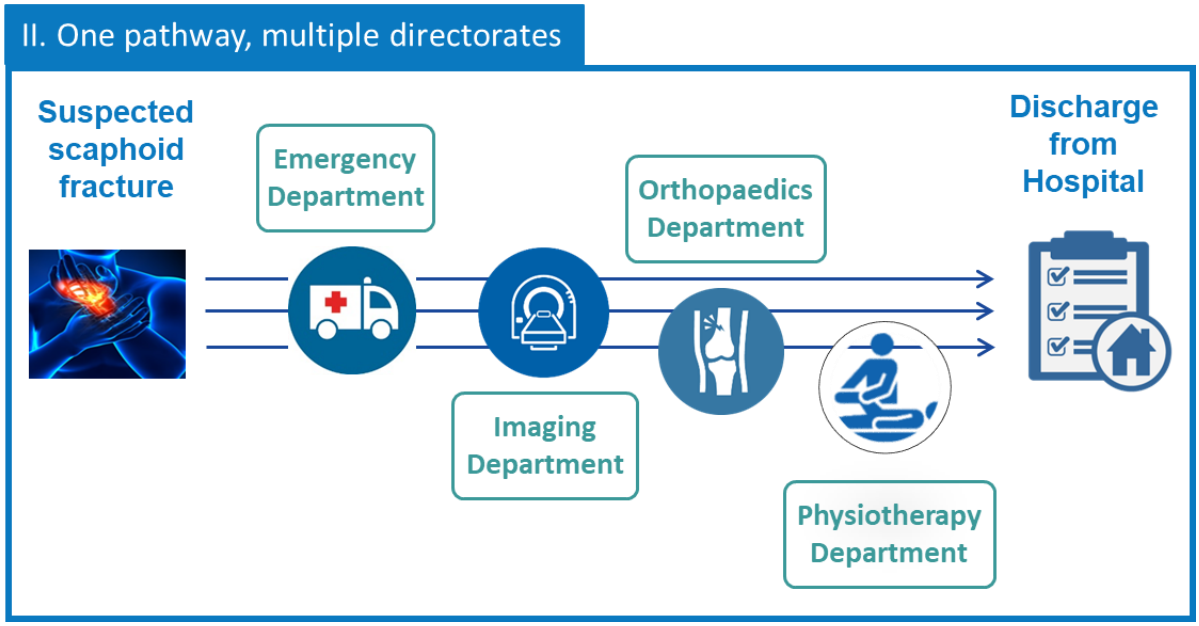


Figure 73. Illustration of the suspected scaphoid fracture crossing multiple clinical directorates.

Although the trial produced favourable clinical and economic results, there are different considerations that posed difficulties to the adoption of immediate MRI in routine practice care. These difficulties were multi-factorial and grouped into four dimensions (financial, organisational, technological and cultural) as illustrated in Figure 74.

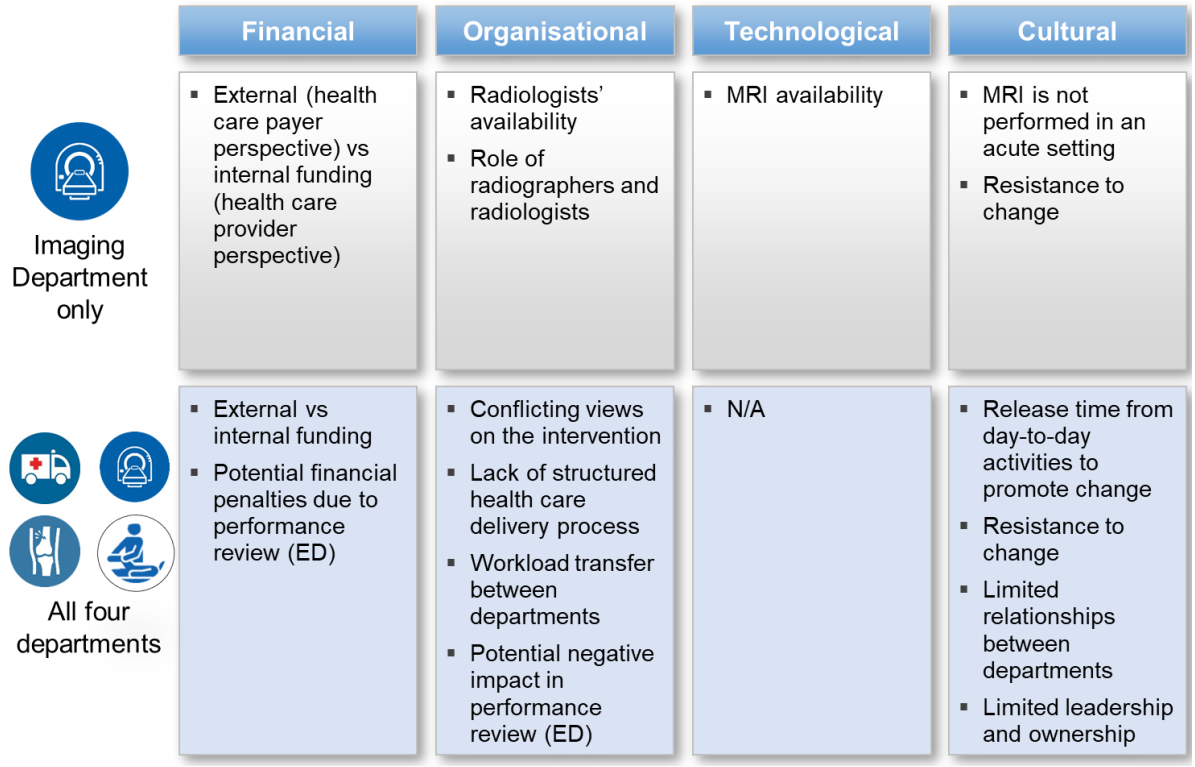


Figure 74. Implementation challenges per dimension of analysis for all four departments considered.

Financial dimension. There were challenges associated with the internal cascade of financial incentives throughout the healthcare provider's organisation. Although the intervention led to cost savings from the healthcare payer's perspective (NHS England), existing funding arrangements might prevent its adoption as the healthcare provider might incur a financial loss by doing so. Furthermore, even if the healthcare provider also benefited financially from the intervention (as is the case of immediate MRI), different departments might incur financial gains or losses as a consequence of adopting the intervention. From the imaging department perspective, there would be an increase in the number of MRIs, thus contributing to a growth in the level of funding. However, the imaging department will charge these MRIs internally to the department of the referrer, i.e. the emergency department. These charges are internal and might not be associated with the income generated by the intervention. In practice, although the income generated by the Emergency Department (ED) increases, this growth is only a fraction of the actual cost incurred by the department. In other words, compared to routine clinical practice, the ED makes a loss for each immediate MRI requested to manage suspected scaphoid fractures. To make matters worse, the introduction of MRI was associated with an increase in the proportion of breaches of the NHS 4-hour ED (21% vs 3.1% in the standard care group) and potential subsequent negative financial incentives. Similarly, the intervention with immediate MRI, reduces the number of outpatient appointments at Fracture Clinic (Orthopaedics department). Even though this could lead to a reduction in income generated by the Orthopaedics department because

outpatient appointments are reimbursed based on activity-based fees (under a contract block arrangement), this was not an issue due to existing waiting lists for appointments.

Organisational dimension. Various organisational challenges arose while implementing the intervention. From the imaging department's perspective, these comprised the limited availability of radiologists to support referrers from the ED with 'live' reporting of wrist MRI scans and the inherent change in the role of both radiographers and radiologists. All four departments (Emergency, Imaging, Orthopaedics and Physiotherapy departments) faced multiple challenges. These ranged from conflicting views on the intervention (e.g. patient selection criteria), the lack of a defined healthcare delivery process agreed by all stakeholders, the potential negative impact in performance targets (4-hour ED target) and workload transfer between departments. The last point was of particular relevance as the intervention led to an increase in the workload for the imaging and emergency departments whilst contributing to decreasing the orthopaedics and physiotherapy departments' workload.

Technological dimension. The main challenge was the availability of MRI slots to accommodate an urgent wrist MRI referral. Given the high utilisation rates of MRI scanners, the appropriateness of immediately imaging patients with suspected scaphoid fracture was questioned by different clinicians.

Cultural dimension. As with any change, cultural challenges arouse. From an imaging department point of view, the intervention fundamentally affected the paradigm around the utilisation of MRI, moving from a non-acute to an acute setting. This led to multiple staff members expressing opposition to this change. From a wider perspective, challenges were related to the limited leadership and ownership of the change process (e.g. "this is a radiology-driven change"), the limited network of working relationships between the four departments and the lack of time for different stakeholders to take time off clinical duties to implement changes across their clinical pathways (continuous change as part of the business as usual practice).

Implementation plan:

The PhD student, in partnership with other members of GSTT, developed an implementation plan to address the challenges previously identified, which comprised the following milestones (Figure 75). Appendix VII details the implementation plan based on the Consolidated Framework for Implementation Research (CRIF) framework previously described in Chapter 2.

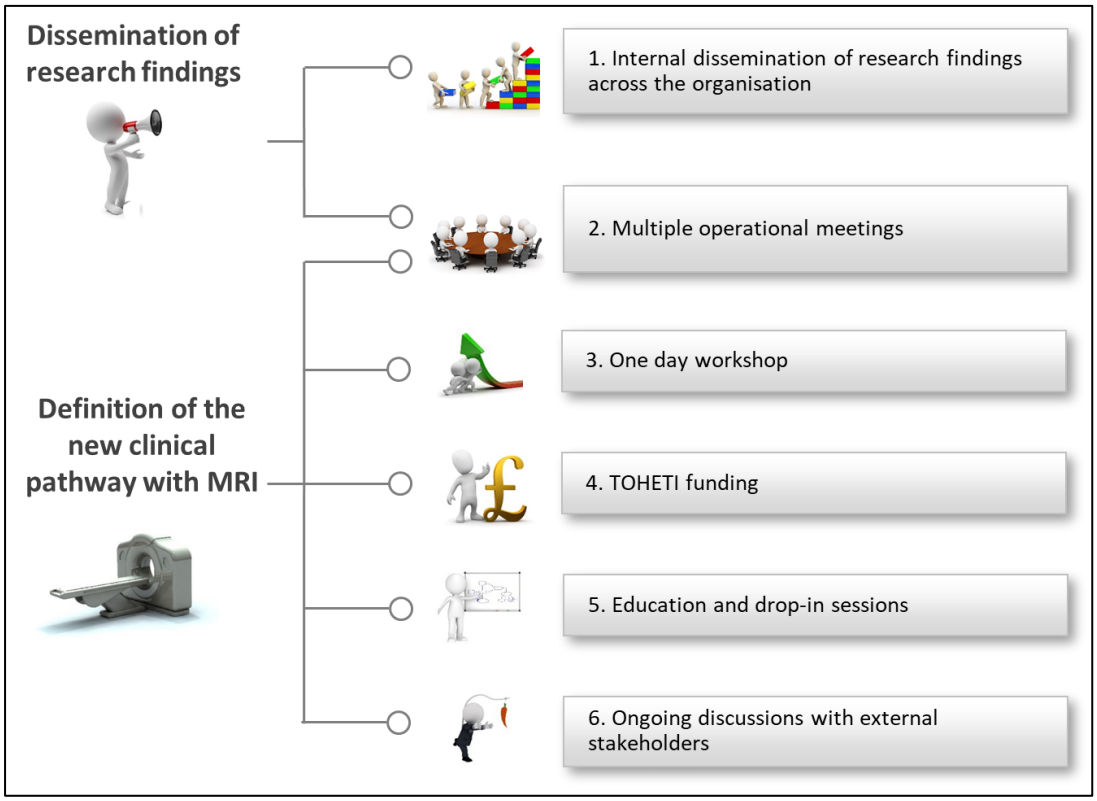


Figure 75. Illustration of the aims and individual elements in the implementation plan.

1. **Internal dissemination of the research findings** across the Trust (via newsletter and internal email from the executive team). In addition, research findings from the SMaRT trial were presented at different directorate clinical governance meetings across all four departments. This approach aimed not only at disseminating the evidence generated but also to give the opportunity for different members of staff to ask questions and actively take part in the process of redesigning the pathway.
2. Multiple **operational meetings** involving individual departments and all four departments combined. These meetings provided a clearer understanding of the intervention's implications and delineated a clear pathway with all clinical and administrative processes associated with its real-world delivery (e.g. who is eligible for the pathway, what to do when eligible patients present out of MRI working hours). In addition, a **workflow protocol** was put in place to avoid breaches associated with the 4-hour ED target. Furthermore, to endorse the priority of the implementation project, senior engagement was involved in the implementation plan. As an example, the GSTT's Chief Executive Officer and Medical Director attended multiple ED huddles at 8am.
3. A **one-day workshop** concerning the roll-out of the new clinical pathway. This event involved clinicians, managers and analysts from all four directorates as well as members of the executive team of the hospital. As part of this event, multiple challenges were discussed and addressed. As an example, the chief executive's support for the implementation of the intervention, despite its internal financial implications, was a major contributing element to its successful adoption. Another

by-product of this top-down approach was to motivate and empower key stakeholders to lead the change.

4. The TOHETI **funding** allowed for the purchase of additional scanners to increase MRI capacity, thus allowing the intervention to be implemented. This was an important element to gain the full buy-in from the imaging department.

5. Ongoing **discussions with external stakeholders**, in particular Clinical Commissioning Groups (CCGs) at Lambeth and Southwark. This aimed to provide CCGs with evidence on the need to revisit existing funding schemes associated with the provision of care for patients with suspected scaphoid fracture (e.g. new tariff that incentivises the use of immediate MRI in the acute setting).

6. **Education and drop-in sessions** were held where members of staff could express their concerns and ideas for the two weeks following the adoption of the new clinical pathway. This provided staff members with a feedback loop and the ability to flag issues with the provision of the intervention.

The timeline with the different milestones of the scaphoid project, from the research component to the actual implementation in routine clinical practice, is illustrated in Figure 76. The project started back in 2013 with initial brainstorming of ideas and identification of the intervention to be evaluated. The following four years comprised three studies, a modelling study, a feasibility study and a pragmatic randomised controlled trial (SMaRT trial). The last year of the project was focused in the implementation of the intervention in routine clinical practice (achieved on Monday 28th October 2019).

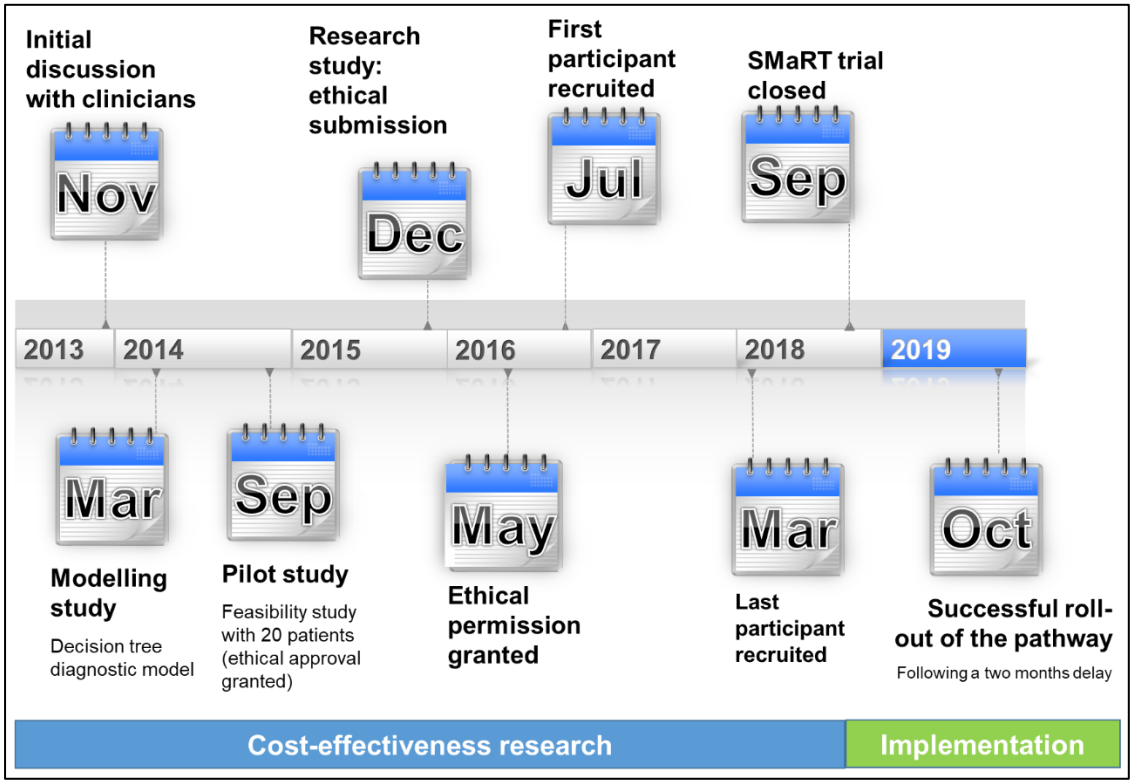


Figure 76. Timeline associated with the project, from its inception, to its research component and the implementation of the intervention as part of routine clinical practice.

Implications for clinical practice at GSTT:

The SMaRT trial changed clinical practice at a major London teaching hospital NHS Trust, leading to the inclusion of immediate MRI in the ED for the management of suspected scaphoid fractures. This meant that care management of over 500 patients annually will potentially be affected as a direct consequence of the research conducted as part of this PhD. In addition, and to our best knowledge, this is the first NHS clinical pathway that uses MRI during the initial ED episode.

7.3.2 Chronic headache

Model-based evaluation:

A 12-month time horizon decision tree model compared the condition-specific healthcare costs of two coexisting clinical pathways in the management of chronic headache that differed in the referral from primary care to either neurology services or direct access to brain MRI. The model considered was however conceptually different from those developed for the scaphoid and colon cancer pathways. The inherent rationale for direct referral for brain MRI for patients with chronic headache was that a negative scan would reassure patients that no underlying life threating condition is causing the disease. Given the very low incidence of such life threatening conditions (<0.1% for brain cancer), this meant that the added value of brain MRI was not driven by its diagnostic accuracy for a given condition but for its potential ability to decrease subsequent use of NHS resources. Furthermore, based on clinical experts, the economic model also incorporated a branch around the diagnosis and management of incidental findings associated with the brain MRI.

The base case scenario estimated that the management of chronic headache with direct referral to brain MRI produced marginal 12-month cost savings (£16) compared to the referral to neurology services. Given the level of uncertainty associated with the cost implications of both management strategies and at-risk population, a prospective, observational study was conducted at GSTT.

Study-based evaluation:

Primary research hypothesis. What are the economic and clinical benefits of using direct referral from primary care to head MRI in the management of patients with chronic headache?

Chapter 4 presented the clinical and economic evidence of two existing clinical pathways for the management of patients with chronic headache. Despite almost entirely being managed within primary care, the increase in referrals to secondary care due to chronic headache has been adding extra pressure to resource constrained hospital neurology services. Within the NHS policy of providing GPs with direct access to imaging, direct access to MRI for chronic headache patients coexisted with the

conventional referral route, i.e. referral to a neurologist. Given the very large population at risk and the potential clinical and economic impact, existing evaluation frameworks, such as Gazelle et al. (2011), recommend the use of high-level evidence. Given that both pathways were established as standard care and, in effect, the referral patterns are selected by the referrer (i.e. the GP), a prospective, observational study was designed to capture differences in terms of costs at 6 months post-recruitment (i.e. either after initial neurology appointment or the MRI scan) from a NHS perspective.

This study found that direct access to MRI led to a significant decrease in the total NHS costs both at 6 and 12 months post-recruitment (Table 88). Given the non-randomised design of the study, the generalised linear model analyses were refitted to adjust for baseline factors for potential imbalance between groups. The adjusted analyses showed that direct access to MRI remained cost saving at both 6 and 12 months post-recruitment.

Table 88. Primary null hypothesis considered in the headache study.

Primary Null Hypothesis	**Null hypothesis rejected / not rejected**
There is no difference in 6-month NHS cost per patient between the use of direct referral from the GP to the MRI services compared to referral from the GP to neurology services.	Null hypothesis rejected. There was a significant difference in total costs between both groups (-£333, 95% CI: - £413 to -£253, p<0.001).

In terms of self-reported quality of life, headache-specific and generic questionnaires were used. Unfortunately, due to very high attrition rates, changes in self-reported quality of life over the study's timeline need to be interpreted with caution. At baseline, participants in the Neurology group reported lower quality of life and a higher headache burden.

Due to the limited utility data retrieved from the EQ-5D-5L questionnaires and the imbalances at baseline, the cost-utility analyses performed need to be interpreted with caution. Despite the lower cost per participant in the MRI group, direct access to MRI does not seem to be cost-effective at conventional willingness-to-pay thresholds. This was due to the intervention with direct access MRI generating a lower number of QALYs compared to referral to neurology services. However, as mentioned in the paragraph above, these results need to be confirmed with further research before any policy change is to be made.

With regards to accessibility, GP direct access to brain MRI led to an improvement, considering the time elapsed from referral to first appointment in both groups, i.e. either the initial neurology appointment or the MRI scan. Even if the MRI report, rather than the actually MRI scan, is considered as the first appointment for reassurance purposes, the intervention with brain MRI still improved access to care compared to referral to neurology services.

Implementation work:

The headache study evaluated two existing clinical pathways that GPs in Lewisham and Southwark use to manage patients with chronic headache. A qualitative study interviewed 20 GPs from these two boroughs and presented two main conclusions (Underwood, Kilner, and Ridsdale 2017). First, that GPs had variable awareness of the direct access to MRI pathway, ranging from some GPs that commonly referred patients to others that were not aware of this management option. Second, GPs mentioned they had difficulty managing incidental findings from the MRI scan and that reports were not appropriate to support their subsequent headache management. These two elements were commonly given as the reasons why the uptake of direct access for MRI for these patients had been slow.

The implementation plan considered two important tasks to tackle the identified challenges (Figure 77 and, for further detail, the CRIF framework on Appendix VII). First, in order to increase awareness across GPs (i.e. the referrers), a GP with special interest in headache management and a neurologist provided training sessions across multiple practices, and also conducted a training workshop with an attendance of over 75 GPs. Second, to proactively support GPs in the management of patients following an MRI scan, the structure and content of the brain MRI reports needed to better reflect the information requirements of the referrer. The latter was particular important if the brain MRI showed incidental findings as GPs commonly stated that they struggled to interpret and manage these findings. This meant that radiologists were expected to not only report the MRI findings but also support their colleagues in primary care with the subsequent management of the chronic headache. For this purpose, the student worked with a consultant radiologist and a consultant neurologist to suggest an amendment to the existing reporting template. This task (ongoing at the time of the writing this chapter) consisted of adding a full section on how the GP should manage the patient's chronic headache on the back of the MRI findings.

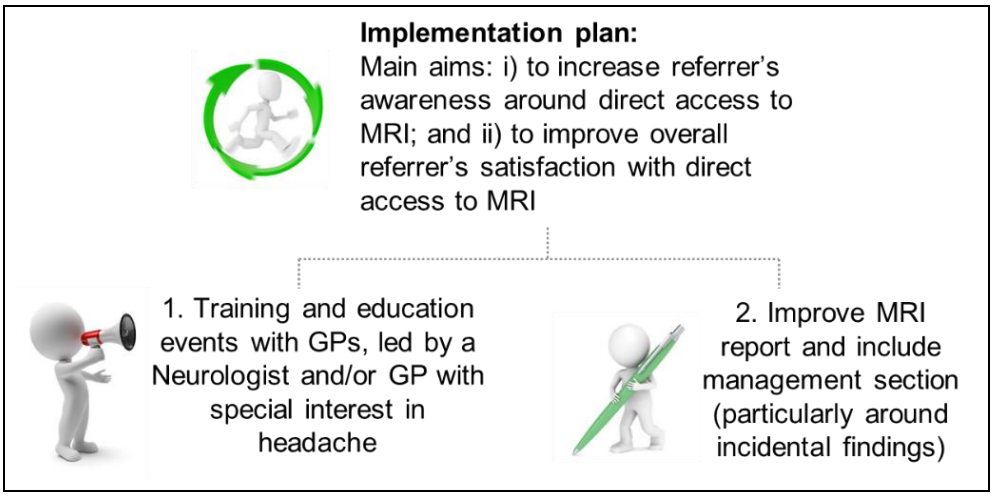


Figure 77. Two major tasks considered in the chronic headache implementation plan.

Implications for clinical practice at GSTT:

The implementation plan aimed to empower and reassure GPs (i.e. the referrers) and ultimately increase the proportion of patients with chronic headache directly referred for a brain MRI. This was corroborated with overall head MRI utilisation at GSTT for all types of referrals, which has increased over 120% in four years, with 465 head MRI scans in 2015, compared to over 1,000 scans in 2019. In just three years (2017 to 2019), the increase in direct referrals to MRI from GPs has more than doubled. As a by-product of this increase, direct access to brain MRI should not only contribute to the NHS financial sustainability agenda but also release neurology capacity for other neurological conditions. However, this increase was still not enough to offset the overall increase in referrals to secondary care due to chronic headache. In other words, although direct access to MRI has diverted some patients from the neurology department, the sheer increase in referrals from primary to secondary care has also led to an increase in referrals to neurology.

7.3.3 Suspected colorectal cancer

Model-based evaluation:

A 6-month time horizon decision tree model compared the condition-specific healthcare costs of Computed Tomography Colonography (CTC) as a direct alternative to optical colonoscopy (OC) in patients with low to intermediate risk of Colorectal Cancer (CRC). The underlying rationale was that using a non-invasive and less costly diagnostic scan (CTC) would, in a large proportion of patients, avoid the need for patients to undergo an invasive, more expensive test (OC). This could ultimately lead to cost savings for the healthcare payer.

The base case scenario estimated that the use of CTC, replacing OC as the initial colonic investigation, led to a £267 cost saving for the NHS. This cost difference was driven by four components: (i) the prevalence of CRC among patients deemed to be at low to intermediate risk of CRC; (ii) the diagnostic accuracy of CTC compared to OC; (iii) the number of extracolonic incidental findings found on CTC; and (iv) the unit cost associated with the provision of both diagnostic tests. When subjected to extensive deterministic sensitivity analyses, CTC remained a cost saving strategy. Nevertheless, given the lack of observed data on the management of extracolonic findings, and the potential clinical implications of missing CRC in a large cohort of patients, some clinicians expressed reservations about using CTC as a direct alternative to OC without local real-world evidence. For these reasons, a prospective, observational study was conducted at GSTT.

Study-based evaluation:

Primary research hypothesis. What are the economic and clinical benefits of using CTC in the management of patients with low risk suspicion of colorectal cancer?

The Department of Health forecasted an annual increase of 10% to 15% in the demand for endoscopies, putting an additional burden on the already overstretched OC capacity (Department of Health 2012b). Backed by evidence of non-inferiority of CTC in the diagnosis of medium to large polyps and CRC, the utilisation of non-invasive CTC as a direct alternative to the gold standard OC was evaluated. Chapter 5 presented the clinical and economic evidence of two existing clinical pathways for the management of patients with suspected colorectal cancer (CRC) that varied in the initial diagnostic scan used, either optical colonoscopy (OC) or Computed Tomography Colonography (CTC). The intervention with CTC was found to be cost saving to the NHS at 6 months post-recruitment (Table 89). Given the non-randomised design of the study, generalised linear model analyses were adjusted for baseline factors for potential imbalance between groups. The adjusted analyses showed that using CTC as the first imaging scan for patients with low to intermediate risk of CRC remained cost saving. In terms of cost-effectiveness, the intervention with CTC had a 91.4% probability of being cost-effective compared to OC at month 6 post-recruitment at the £20,000 willingness-to-pay per QALY threshold.

Table 89. Primary null hypothesis considered in the colon study.

Primary Null Hypothesis	Null hypothesis rejected / not rejected
There is no difference in 6-month NHS cost per patient between the use CTC as the first line investigation compared to OC for patients with low to intermediate risk of CRC.	Null hypothesis rejected. There was a significant difference in total costs per participant between both groups (-£345, 95% CI: -£501 to -£190, p<0.001).

Implementation work:

Two main challenges were associated with the uptake of CTC in the diagnosis of patients with suspected CRC. First, there was an existing perception among clinicians (particularly the referrers) that the CTC pathway presented issues, particularly around the bowel preparation prior to the procedure. Second, capacity issues led to the limited availability of CTC. Faced with these challenges, the referrers (i.e. clinicians running the virtual telephone clinic for all patients referred from primary care with suspected colorectal cancer) historically opted to mostly direct patients to CTC if they were not fit for a colonoscopy.

The implementation plan considered three tasks to tackle the identified challenges (Figure 78 and for further detail, the CRIF framework in Appendix VII).

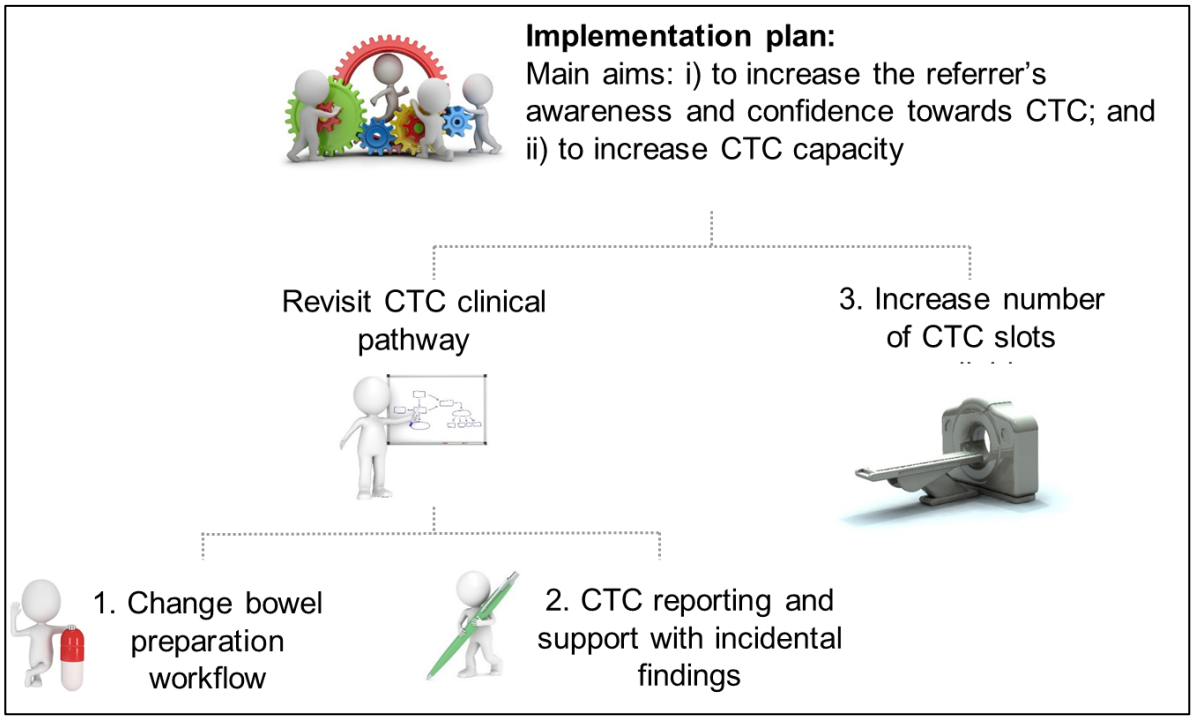


Figure 78. Three major tasks considered in the colon cancer implementation plan.

The first two tasks aimed to address the low awareness and improve the referrer's confidence around the CTC pathway. For this purpose, the entire CTC clinical pathway, from referral to the report being available to the referrer, was fully assessed, with particular focus on areas requiring improvement. A two-week clinical audit conducted by the student confirmed that over three quarters of issues with the pathway resulted from the pre-test bowel preparation. This in turn contributed to an overall 10% CTC cancellation rates. In order to minimise this problem, the way the bowel preparation was given to patients was profoundly changed, moving from a system where the patients physically attended GSTT pharmacy towards an alternative where the bowel preparation and respective patient information sheet was posted to the patient. In parallel, a simpler patient information sheet and substitute bowel preparation drugs with easier to follow instructions (despite being more expensive) were put in place to minimise the risk of non-compliance. In order to further streamline the pathway, revised electronic forms were put in place for the referrer to request the scan and for the radiology department to vet referrals and sign the prescription for pharmacy to post the bowel preparation. The rationale for these changes was to avoid unnecessary patient visits to the hospital and moreover avoid patients turning up on the day without bowel preparation (on the day cancellation) or patients with incomplete or suboptimal colon diagnostic scans (leading to a repeat test).

Second, following the colonic investigation, the CTC report was revised to include a section on incidental findings, particularly extra-colonic findings. This aimed to support the referrer in terms of management advice and to ultimately increase their levels of confidence with the CTC pathway. In

order to disseminate these changes among referrers, multiple training sessions were organised between the radiology and gastroenterology departments.

Third, in order to enhance the responsiveness of the CTC pathway to patients on a two-week wait pathway (suspected colorectal cancer), the number of CTC slots were doubled. In contrast to previous scheduled arrangements, these slots were ring-fenced until two days before the actual scan. This aimed to protect the CT's availability for CTC scans whilst avoiding those CT slots not being utilised.

Implications for clinical practice at GSTT:

The CTC utilisation rate has expanded by 186% in just four years, from 607 CTC scans in 2015 to 1,733 in 2019 (Figure 79). Although these numbers included all patients, regardless of their referral criteria, the intervention with CTC for low to intermediate risk patients was responsible for more than two thirds of this increase. During the same period of time, the number of optical colonoscopies also increased, however at a more moderate rate (over 5 to 10% annually between 2015 and 2019). Despite the use of CTC as a direct replacement test decreasing the need for OC (workload shift from OC to CTC), the sheer increase in referrals for colonic investigations has offset any potential decrease in the number of OC scans. Based on the results from the colon cancer study and in order to accommodate the expected growth in demand, the imaging department has plans to double the capacity of the existing CTC service over the next 2 to 3 years.

The second implication from the implementation plan was the reduction of the on the day cancellation and did not attend (DNA) rates. A reduction of 56% in CTC cancellation or DNA rate was verified, from a 16% overall rate in 2015 to 7% in 2019 (Figure 79). The improvement in the workflow and easier to follow bowel preparation instructions led to a higher proportion of optimal colonic investigations, thus reducing on the day CTC cancellation rates and the need for repeated colonic tests.

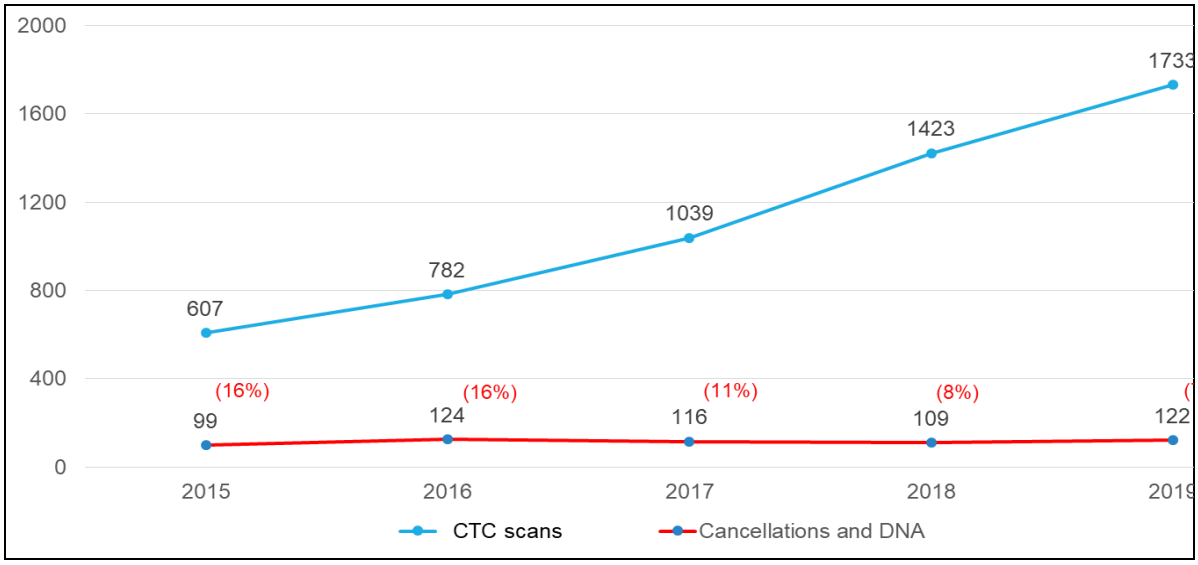


Figure 79. Number of CTC scans and respective cancellation and DNA episodes.

7.4 Strengths and limitations

The strengths and limitations of each studies were discussed in the individual chapters (chapters 3, 4 and 5). This chapter focusses on the overall strengths and weaknesses of the PhD and the research programme TOHETI on which the PhD was based.

Strengths:

The greatest strength of the TOHETI programme was its originality, in aiming to transform healthcare using the new or novel utilisation of advanced imaging as the driver for change. This constituted a paradigm shift as advanced imaging, due to its higher costs and lower availability compared to basic imaging (e.g. radiographs), is usually reserved for later, not early stages of diagnostic pathways. Although this rationale might have been appropriate in the past, the decrease in the unit cost of advanced imaging modalities and their increased availability meant that the proposed interventions led to improved diagnostic and, where needed, treatment pathways, whilst generating cost savings in multiple clinical settings.

Potential initiatives were selected based on a bottom-up approach, i.e. the initiatives were recommended by the clinical teams responsible for the delivery of care. This process empowered key clinicians whilst providing a research framework to assess the holistic use of medical imaging. Furthermore, literature reviews and critical appraisal of published clinical and economic evidence provided a solid foundation on which to base the interventions considered. Given the central role of medical imaging across different clinical pathways, the innovative use of advanced imaging not only impacted that section of the clinical pathway but the sections of pathway before and after the imaging test. As an example, the inclusion of immediate MRI in the management of suspected scaphoid fractures fundamentally changed the subsequent diagnostic and treatment pathway. This meant that, although the intervention took place in the Radiology department, its impact was felt prior to the imaging test, in the Emergency Department, and following the imaging test, in the Orthopaedic Department. This holistic assessment of the pathway, from referral to discharge, regardless of the location of the healthcare resources utilised, was a major strength associated with all studies performed as part of this research programme.

Linked to the process of selecting different initiatives, the involvement of multidisciplinary teams and expertise (e.g. clinical experts, hospital managers, statisticians, health economists) from the onset of the research programme also constituted a major strength of the research programme. This multidisciplinary engagement, particularly during the study design process, was essential to evaluate all the anticipated clinical and health economic implications of each of the initiatives. Key decisions included the pragmatic evaluation in the context of real-world patients and the need to conduct any analysis based on intention-to-treat principles. The aim was to provide health policy makers and commissioners with appropriate evidence on which to base their decision.

Other strengths of the research programme derived from the methodology considered in the data collection and data analysis processes. First, the *a priori* power calculations used in the sample size estimate were performed according to statistical principles and overseen by a Professor of Medical Statistics. Second, all data collection methods were designed to comprehensively capture relevant clinical and economic data during an appropriate follow-up period (dependent on the clinical condition being evaluated). Healthcare utilisation was captured using multiple databases from secondary care (up to seven clinical and financial databases) at an individual level (patient by patient). In addition, based on the holistic approach mentioned, data from primary care was also individually retrieved to obtain healthcare utilisation within primary care (e.g. GP appointments) as well as any other resource utilisation in other hospitals other than GSTT. A third layer of data were retrieved from self-reported participant diaries collected quarterly. Although very time consuming, this comprehensive methodology enabled the use of a large dataset of participants (over 500 patients) with minimal missing resource use data and ensured that all relevant costs were included in the economic analyses. The study design and methods considered aimed to minimise biases and optimise data completeness and integrity.

Contrary to the majority of published evidence, the inclusion of healthcare utilisation derived from incidental findings was considered. As an example, if an incidental finding was diagnosed in the advanced imaging test, i.e. a finding unrelated to the referral or clinical condition, its clinical and cost implications were included in the analysis. This rationale was deemed appropriate as the incidental finding was a direct consequence of the intervention proposed and its exclusion might have led to different results.

The clinical staff responsible for the delivery of care were included as members of the research team. This was an important strength and essential to maximise recruitment rates across the three studies. Given the different number of studies simultaneously run by the PhD student at different physical locations, it was essential to train staff members for them to become independent in the recruitment of participants.

One final strength of the research programme was its funder, a not for profit organisation (Guy's and St Thomas' Charity). The PhD student was fortunate to be involved from the onset on a £13 million grant that allowed not only the purchase of equipment (MRI and CT scanners) but also to evaluate the implications of using those scanners in the context of real-world patients and pathways. The PhD student was given intellectual independence to impartially evaluate the different interventions considered.

Limitations:

In addition to the limitations discussed in each of the individual study chapters, an important weakness derived from the single-site design (one NHS Trust) of all three research studies. This raises potential generalisability issues of the findings reported in each chapter. Although the interventions were

designed to recruit across a heterogeneous population reflecting UK clinical practice, this remains as one important weakness. Prior to widespread adoption of the interventions evaluated further research in different settings and healthcare providers should be considered. The updated models could also be used by decision makers in different health care settings.

All studies were powered on cost differences between groups rather than clinical or cost-effectiveness differences. However necessary for feasibility purposes, it constituted a limitation of the research programme. In order to mitigate this, secondary outcomes considered the clinical and economic evaluation of the intervention. All three interventions were also evaluated based on the incremental cost per QALY as per NICE recommendations.

Lastly, an important limitation in one of the observational studies (chronic headache study) was due to the very high follow-up attrition rates of participants' self-reported data (e.g. utility data). Although financial incentives and non-traditional communication methods (e.g. phone app) were put in place to maximise participant compliance, the difficulty to gather high-quality information from participants was underestimated. This meant that the cost-utility analysis performed lacked robust data and therefore only the cost analysis (and not the cost-utility analysis) could be used to inform healthcare policy makers. For future studies, the study design should consider this issue by prioritising follow-up workload (e.g. using one instead of two headache-specific questionnaires) and consider the resources necessary to intensify follow-up contacts.

7.5 Implications for policy and clinical practice

The implications for policy and clinical practice were discussed in the three previous sections. This section discusses the national implications of using advanced imaging in new or novel ways.

Value-based healthcare principles aim to provide policy makers with appropriate evidence on which to base their decision. The inclusion of different dimensions of analysis, from efficiency (e.g. cost per case, cost per QALY), accessibility to care (e.g. time from referral to diagnosis or treatment), clinical outcomes (e.g. diagnostic accuracy, proportion of fractures healed) to patient satisfaction (e.g. self-reported quality of life) aimed to capture meaningful and tangible outcomes while respecting the physical and psychosocial needs and expectations of our patients.

Irrespective of the intervention investigated, this PhD is aligned with existing healthcare policies in three key areas. First, all interventions considered in the TOHETI programme aimed at improving access to medical imaging. Although only one of the cancer projects was presented in this thesis, 50% of the TOHETI projects (3 out of 6) aimed at improving cancer outcomes by promoting access to medical imaging (NHS Improvement Diagnostics 2012). Second, all interventions promoted both horizontal and vertical integration of care. For instance, the use of immediate MRI in the management of suspected scaphoid fractures promoted the coordination of care across a continuum of care within

secondary care, in this case four different directorates. This transformative work had the potential to not only improve care for patients with suspected scaphoid fracture but also other patients that flow through these departments. Direct access to brain MRI from primary care promoted vertical integration of care from primary and secondary/tertiary care. Third, the interventions endorsed the NHS RightCare initiative under the "*right test, right time, right patient*" principles. As an example, the use of immediate MRI in the management of suspected scaphoid fractures encouraged of the use of a more accurate test (right test) for patients with clinical suspicion of scaphoid fracture and negative radiographic evidence (right patient) during the initial acute episode (right time).

Finally, if advanced imaging is to play a central role in the transformation of healthcare, particular attention needs to be given to NHS capacity. As summarised in Chapter 2 and further detailed in Cake, Cavanagh, and Gordon (2015) and Bainbridge (2018), medical imaging demand is expected to keep increasing due to multiple factors such as the lowering of referral thresholds, inclusion of imaging in new pathways, among others. However, in order to accommodate this increase and shift in demand, considerations about capacity, of both equipment and personnel, are essential. Even with increased efficiency or longer opening hours, the existing capacity within the NHS does not seem to have enough operational flexibility to meet this rise in demand without further investment. Bainbridge (2018) produced a report highlighting that CCGs seem to be aware of this issue, but limited and contradicting evidence around the expansion of diagnostic capacity was presented. Hence, in order to implement change and transform care using medical imaging as the driver for change, policy makers should address regional disparities in imaging capacity and close the gap in terms of number of scanners per inhabitants compared to other developed healthcare systems (e.g. France, Germany).

7.6 Implications for further research

A number of implications for further research were previously described as part of the individual studies (chapters 3, 4 and 5) and the methodological chapter (chapter 6).

Two common features were associated with study attrition rates and statistical methods used in the cost analysis. First, attrition rates varied based on several parameters such as the follow-up period, clinical condition and intervention being evaluated. Longer follow-up periods had higher attrition rates. Clinical conditions with higher disease burden seemed more likely to remain engaged in the study compared to healthy participants or participants with a lower disease burden. Similarly, interventions likely to provide reassurance to participants seems to have led to participant disengagement. For instance, in the evaluation of direct access to MRI for patients with chronic headache, a higher attrition rate in the MRI group was recorded compared to the neurology group. Second, further research into the methodologies used in the cost analyses and the potential impact on the recommendations should be considered. Contrary to most published economic literature, GLMs utilising the identity link function instead of a log link should be considered in order to avoid transformation methods and respective

potential biases (Polgreen and Brooks 2012; Peacock, Kerry, and Balise 2017). Also, the use of bootstrap techniques, commonly used to perform cost-effectiveness and cost-utility analyses, should be further evaluated and compared against GLMs. Although some differences in the 95% confidence intervals for the cost difference between groups using GLM and bootstrap methods were found, the recommendations for the adoption of any of the three interventions would not have changed. Nevertheless, in theory, different policy recommendations could be reached using different statistical methods and therefore further research is required to establish a standard statistical method in the analysis of cost variables.

Another recommendation for further research is to address the rise in demand for advanced medical imaging. Besides capital investment in equipment, workforce redesign is essential to address future challenges. The development of a workforce redesign framework in the context of real-world delivery of care should be considered in the translational research field. Two particular topics were identified as part of the TOHETI programme. First, the current workforce roles should be redesigned to adapt to the ever changing needs of care delivery. For instance, the SMaRT trial evaluated the use of radiographers (technician responsible for the acquisition of images) as a first-line reporter to rule-in or rule-out fractures in the ED. The higher availability of radiographers, compared to radiologists, makes them an ideal resource to expedite this time sensitive acute pathway. This would not preclude the need for a subsequent MRI report from the radiologist but would operationalise the intervention in NHS Trusts without or with limited on call radiologists or musculoskeletal radiologists. Supplementary research that assesses intra and inter-variability of radiographers compared to radiologists (assumed to be the gold standard) should be considered.

Second, the use of artificial intelligence and computer assisted diagnosis will fundamentally change different workflows associated with the reporting of imaging sets. These developments are likely to affect all professionals, particularly radiologists. The role of radiologists is likely to evolve, moving towards reporting of complex imaging scans where computer aided technologies have less traction, interventional radiology and assuming a more proactive role as care coordinator. Future radiology reports should be geared towards not only the diagnosis and reporting of the imaging findings but also to support the referrer with clear management plans. This subject however is deemed contentious and further qualitative and quantitative research is required to assess both the feasibility and the views of radiologists towards the expansion of their clinical role.

7.7 Conclusion

The clinical and economic impact of three imaging-based interventions were evaluated based on health economic and value-based health principles, contributing to new knowledge across these clinical conditions. Based on the study findings, the use of immediate MRI in the ED was added to standard care in the management of suspected scaphoid fractures. In addition, two existing imaging-based

pathways for management of chronic headache and suspected colorectal patients were promoted as alternatives for selected patients. In all, these three interventions hold the potential to improve the clinical care of over 3,000 patients annually at GSTT.

Despite its limitations, this research programme challenged the broadly accepted paradigm that advanced imaging should be reserved for later stages of the diagnostic pathway. This new imaging paradigm aims to promote the early use of accurate and definitive diagnostic tools, streamlining the diagnostic and, if needed, treatment pathway. This will ultimately lead to improvements in both clinical and patient outcomes whilst reducing overall costs of care.

Although intangible or difficult to quantify, a second consequence of this research programme was the inherent cultural change across different directorates of one organisation as important as GSTT (over 15,000 employees). Conscious of the added value of the PhD and the TOHETI programme, the hospital executive board decided to incorporate its approach and methodology as part of a novel organisational wide transformation programme, called Care Redesign. In many ways, this is one of the main legacies of this PhD and the TOHETI programme.

Appendices

Appendix I. Systematic literature review search strategy

Database: Embase <1974 to 2018 Week 48>, Ovid MEDLINE(R) and Epub Ahead of Print, In-Process & Other Non-Indexed Citations and Daily <1946 to November 26, 2018>

Search Strategy:

---

1 phased approach.ab. (452)

2 hierarchical approach.ab. (1101)

3 phased evaluation.ab. (13)

4 hierarchical approach.ab. (1101)

5 hierarchical evaluation.ab. (37)

6 framework.ti. (57932)

7 diagnosis.ti. (738539)

8 diagnost*.ti. (318537)

9 imaging.ti. (507324)

10 1 or 2 or 3 or 4 or 5 or 6 (59507)

11 7 or 8 or 9 (1524153)

12 10 and 11 (962)

13 limit 12 to yr="2009 -Current" (631)

14 remove duplicates from 13 (377)

****

Appendix II: Search strategy used in the search of economic evidence around the utilisation of advanced imaging in the management of suspected scaphoid fractures.

Database: Ovid MEDLINE(R) <1946 to May Week 2 2016>, Ovid MEDLINE(R) In-Process & Other Non-Indexed Citations <May 18, 2016>

The search strategy returned **93 papers**.

---

1 (scaphoid adj5 fracture).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier] (917)

2 (suspected adj5 scaphoid).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier] (182)

3 (occult adj5 scaphoid).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier] (79)

4 scaphoid.ti,ab. (3811)

5 1 or 2 or 3 or 4 (3812)

6 exp magnetic resonance imaging/ (347899)

7 MRI.m_titl. (43994)

8 Tomography, X-Ray Computed/ (317812)

9 CT.m_titl. (61578)

10 "Bone and Bones"/ or Bone Neoplasms/ (130052)

11 bone scintigraphy.m_titl. (1813)

12 exp emergencies/ (36558)

13 acute.m_titl. (410280)

14 emergency.m_titl. (66567)

15 advanced imaging.m_titl. (246)

16 (compute* and tomograph*).mp. (478230)

17 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 (1347678)

18 exp radiography/ (687469)

19 exp x-ray/ (26428)

20 (plain adj5 x-ray).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier] (2189)

21 (plain adj5 radiography).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier] (3188)

22 (conventional adj5 radiography).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier] (3092)

23 (scaphoid adj5 radiography).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier] (45)

24 18 or 19 or 20 or 21 or 22 or 23 (714988)

25 costs.mp. and cost analysis/ [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier] (44066)

26 (economic adj3 evaluation$).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier] (8630)

27 (economic adj3 analy*).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier] (6450)

28 (economic adj3 (study or studies)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier] (3400)

29 (cost adj3 evaluation$).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier] (2307)

30 (cost adj3 analy*).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier] (116802)

31 (cost adj3 (study or studies)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier] (7809)

32 (cost adj3 effective*).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier] (96096)

33 (cost adj3 benefit*).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier] (73070)

34 (cost adj3 utilit*).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier] (3808)

35 (cost adj3 minimi*).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier] (2411)

36 (cost adj3 consequence*).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier] (706)

37 (cost adj3 comparison*).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier] (2057)

38 (cost adj3 identificat*).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier] (394)

39 25 or 26 or 27 or 28 or 29 or 30 or 31 or 32 or 33 or 34 or 35 or 36 or 37 or 38 (191984)

40 exp Carpal Bones/ or exp Wrist Injuries/ or exp Scaphoid Bone/ or exp Fractures, Bone/ (163724)

41 5 or 40 (164435)

42 17 and 24 and 39 and 41 (113)

43 limit 42 to yr="1990 -Current" (98)

44 limit 43 to (english or portuguese or spanish) (93)

****

Appendix III: Data extraction from the fifteen papers included in the systematic literature review (grouped as per the study design).

Authors (with date, country)	Study design and follow-up duration	Intervention(s) and comparator(s)	Population characteristics and sample size	Economic data/economic related outcomes / clinical outcomes	Main economic findings	Author's conclusion
Randomised studies						
Kelson et al. (2016) Australia	**Study design:** Randomised controlled pilot. **Follow-up:** At 42 days, using questionnaires.	**Intervention:** Early MRI scan performed 1-3 days from day of presentation in patients with negative radiographic findings in the initial X-ray. **Comparator:** immobilisation with scaphoid plaster and periodic clinical and radiographic follow-up.	19 patients in total, 11 in the MRI group and 8 in the comparator group.	**Costs:** Health care resource use was obtained via a questionnaire at 42 days. The cost of healthcare was derived from Medicare Benefits Schedule 2014. **Clinical outcomes:** pain and function scores, immobilisation time (measured in days). **QALYs:** None. **Economic evaluation:** not applicable (cost analysis only).	**Costs:** The comparator group has a median cost per person of $486.90 (AUD) (149.51–724.63) - [$337.09, $103.51-$501.67], slightly higher when compared with $485.05 (AUD) (448–550.23) - [$335.81, $310.16-$380.94] for the MRI group (p= 0.74). **Other results:** MRI reduces the immobilisation time (2.5 days vs 29.5 days, p=0.026) and treatment time (3 days vs 31 days, p=0.004). Patient satisfaction scores were higher in the MRI group, but the difference was not statistically significant.	The use of early MRI in the diagnosis of suspected scaphoid fractures resulted in less visits to healthcare specialists and appears to be slightly cost-saving for the healthcare perspective.
Patel et al. (2013) UK	**Study design:** Prospective, unblinded, non-parallel group, randomised controlled trial. **Follow-up:** Follow-up period up to 42 days.	**Intervention:** Patients with negative findings in the initial X-ray are treated with a removable scaphoid cast and randomised to the MRI group (MRI scan to be performed within two working days from A&E attendance). Subsequent treatment based on MRI findings (if negative, no follow-up is arranged). **Comparator:** Patients with negative findings in the initial X-ray are immobilised with removable scaphoid cast and randomised to the control group, with clinical follow-up at 14 days.	84 patients were randomised into MRI (45) and control (39) groups. There were no baseline differences apart from greater dominant hand injuries in the MRI group. All patients had negative radiographic evidence of scaphoid fracture at A&E presentation.	**Costs:** Mean cost per patient in the management of both groups (MRI vs X-ray only). Management costs for all patients were calculated using data from the Finance Department at West Middlesex University Hospital (UK). Relevant direct, indirect and overheads costs were included in the cost analysis. Time off work was evaluated but not included in the cost analysis. **Clinical outcomes:** Mean pain and patient satisfaction scores. **QALYs:** None. **Economic evaluation:** not applicable (cost analysis only).	**Costs:** No statistically significant differences were found in terms of mean management costs with £504.13 ($728.80) in the MRI group and £532.87 ($770.35) in the control group (p=0.9). **Other results:** No statistically significant differences were found in terms of better pain and satisfaction scores (MRI was found to have a trend of both better pain and satisfaction scores).	Early MRI in the diagnosis of scaphoid fractures is marginally cost saving compared to conventional management.

Brooks et al. (2005) Australia	**Study design:** Non-blinded, randomised controlled trial. **Follow-up:** period of 3 months (patients contacted monthly to collect information on resource use and recovery of wrist function and pain).	**Intervention:** Patients with negative findings in the initial radiography underwent an MRI scan within two to five days from A&E attendance. Apart from the MRI, no other alterations to the clinical pathway were made. **Comparator:** current clinical practice with immobilisation with scaphoid plaster and periodic clinical and radiographic follow-up.	28 patients were randomised into MRI (11) and control (17) groups.	**Costs:** Costs of health services were retrieved from the Medicare Benefits Schedule (2002). The study seems to assume a societal perspective, including healthcare direct costs as well as indirect costs, associated with days off work. **QALYs:** None **Economic Evaluation**: Cost-effectiveness analysis, using as the measure of effect the number of days patients unnecessarily spent in plaster.	**Costs/Cost-effectiveness:** Derived from the bootstrap simulation, the study estimates a $30.8 per day saving due to prevention of unnecessary immobilisation by the use of MRI (95% CI $2.97 to $69.94). The MRI group used fewer healthcare units (median 3.0, interquartile range 2.0–4.25) than the control group (5.0, 3.0–6.5) (p = 0.03 for the difference). The median cost of healthcare in the MRI group ($411.48, $381.71– $461.93) was slightly higher than in the control group ($296.42, $86.12– $496.46) (p = 0.19 for the difference). At a median productivity loss of 50%, the addition of MRI is likely to be cost-effective 95.3% of the time.	The use of MRI reduces the number of days of unnecessary immobilisation and the use of healthcare units. Health care costs increased non-significantly in relation to the use of MRI in this setting. However, when productivity losses are considered, MRI is likely to be cost-effective.
Quasi-experimental studies						
Bergh et al. (2014) Norway	**Study design:** Pseudo-randomised controlled trial (treatment allocation dependent on the presentation day). **Follow-up:** short-term (not specified). Patient's A&E records assessed after two years to identify potential missed scaphoid fracture.	**Interventions:** MRI following normal radiography, usually performed within mean 1 day (0–7) after attending A&E. Follow-up consultation was booked regardless of the MRI results. **Comparator:** conventional treatment with a below-elbow scaphoid cast for 2 weeks, at which point a follow-up consultation and/or radiography was obtained. If diagnosis is not clear after 2 weeks, patients would undergo an MRI.	124 patients between 18 and 49 years of age who attended Bergen Accident and Emergency Department (Norway) with suspected scaphoid fracture during 1 year in 2009–2010.	**Costs:** Direct costs associated with the diagnosis/treatment - obtained from the Norwegian Health Economics Administration - and indirect costs associated with sick leave retrieved from the Norwegian Labour and Welfare Service. **QALYs:** None. **Economic evaluation:** Cost-minimisation evaluation (cost analysis).	**Costs:** No statistically significant differences in the total direct medical costs between the two groups. However, average direct costs in the MRI group (4,308€ or $4,805) than in the control group (6,999€ or $7,806). Indirect costs for employees (due to sick leave) represented 85% of the total management cost. **Other results:** Patients in the MRI group without a scaphoid fracture (n=54), but potentially with other fractures, used a cast for fewer days (mean 8.8 day) than patients in the control group (n=59) (difference of 6.5 days; p < 0.005). There is a statistically	In a Norwegian setting, the early use of MRI was of value in patients with clinically suspected scaphoid fracture and normal findings in the conventional radiography.

					significant difference between groups in terms of sick leave (7 days vs. 15 days; p = 0.002).	
Moreno-Ramos et al. (2013) Spain	**Study design:** Prospective, non-randomised controlled trial. **Follow-up:** Not specified but includes the full diagnostic pathway until a definitive diagnosis is reached.	**Intervention:** MRI study if the findings on X-ray continued to be negative at the first follow-up examination 10 days after trauma. **Comparator:** Immobilisation and periodic clinical and radiographic follow-up (X-ray and CT in the final phase of the process).	33 cases of patients with clinically suspected fractures of the scaphoid and negative findings on initial X-ray.	**Costs:** Direct and indirect costs were considered - cost per protocol considered in each group. **QALYs:** None **Economic evaluation**: not applicable (cost analysis only).	**Costs:** The cost of the MRI protocol was 131.06€ ($194.58) per patient whilst the cost of the traditional protocol was 114.41€ ($169.86) or 151.06€ ($224.28) per patient, depending on the follow-up studies required.	The cost of the MRI protocol is similar to the comparator and even lower in some cases (patients that required the use of CT during the follow-up period).
Hansen et al. (2009) Denmark	**Study design:** Prospective, non-randomised controlled trial. **Follow-up:** period of 3 months (patients interviewed at 3 months to ask the duration of leave and hand function).	**Intervention:** Patients with negative findings in the initial X-ray underwent an MRI scan within a week from A&E attendance. Subsequent treatment based on MRI findings (if negative, no follow-up was arranged and patient was incentivised to use the hand for daily activities). **Comparator:** Standard treatment with immobilisation with elbow dorsal splint and clinical and radiological follow-up at 2 weeks. If clinically suspicious at 2 weeks, the wrist was splinted again and reviewed after 4 weeks and sometimes after 6 weeks.	27 patients with negative findings in the initial X-ray were assigned to each group using non-random methods (two Danish hospitals, one using the intervention and the second the comparator).	**Costs:** Average of total costs for both groups. The number of hospital contacts were retrieved from case records. All healthcare contacts were costed using Danish tariffs. The study used both a hospital and non-hospital (societal) perspectives. Productivity loss costs were valued using a Danish daily average rate. **Other outcomes:** Time off work and immobilisation time per group (measured in days). **Clinical outcome**: hand function score. **QALYs:** None. **Economic evaluation:** Cost-effectiveness analysis, using the hand function as the measure of effect.	**Costs:** A difference in hospital costs of €151 ($209.89) in favour of the standard treatment observed (p<0.05). A difference in non-hospital costs of €2,869 ($3,988) in favour of the MRI group supported the use of advanced imaging (p<0.05). **Other results:** Immobilisation time reduced from 20 days (range 6–54) in the radiography group to 4 days (range 1–19) in the MRI group (p<0.01). Time off work reduced from 27 days (1–92) in the radiography group to 11 days (0–28) in the MRI group (p<0.01). Median hand function score at follow-up in the radiography group was 2 (1–3), and 2 (1–4) after MRI supported treatment (p=0.70). The adjusted difference was 0.2 (95% CI, -0.3–0.7) in favour of standard treatment (p<0.224).	The use of MRI increased hospital costs (p<0.05), but, at the same time, reduced non-hospital costs by $4,068, due to the reduction of days off work.
Gooding at al. (2004)	**Study design:** Non-randomised pilot study (further details	**Intervention(s)**: Patients with negative findings in the initial radiography underwent an MRI	Patients presenting to an A&E Department	**Costs:** Resource use was collected using data from patients included in the study.	**Costs/Cost-effectiveness:** Average medical cost was NZD 470 ($708) for the standard	The early use of MRI is cost-effective ant it is

New Zealand	on the study design are missing). **Follow-up:** Not specified.	(to be performed within 1-3 days of initial presentation). **Comparator**: Immobilisation with scaphoid plaster and periodic clinical and radiographic review (at 2, 4 and 6 weeks).	in New Zealand with suspected scaphoid fracture and negative X-ray result were allocated to the treatment (50 patients) or the comparator group (40 patients).	Key direct costs were considered – e.g. X-ray, plaster, assessment and MRI - to estimate the total cost per group. **QALYs**: None **Economic evaluation:** Cost-effectiveness analysis, with an ICER using the identification of fracture as measure of effect.	intervention group and NZD 533 ($803) for the MRI group. Average cost to exclude a fracture was NZD 459 ($692) for the standard intervention group and NZD 437 ($659) for the MRI group.	recommended to be offered as part of the routine diagnosis.
Economic evaluation (using economic models)						
Yin et al. (2015) China	**Study design:** Economic modelling using a decision analysis model (decision tree). **Follow-up:** 2 weeks	**Interventions:** 6 interventions were considered: immediate CT, day 3 MRI, day 3 bone scan, week 2 X-ray alone, week 2 X-ray and CT, week 2 X-ray and MRI, week 2 X-ray and bone scan, and immediate MRI **Comparator:** No specific comparator was established (the 6 interventions were compared among themselves).	Patients presenting with suspected scaphoid fracture in an acute care setting.	**Costs:** Costs and benefits calculated from a societal perspective. Resource use and benefits retrieved from literature. Costs were estimated using 2013 Medicare data. **QALYs:** None **Economic evaluation:** cost-effectiveness analysis. ICER per detected scaphoid fracture is estimated.	**Costs/cost-effectiveness:** Immediate imaging (CT or MRI) and day 3 MRI were the most cost-effective strategies for diagnosing suspected scaphoid fractures. ICER of immediate MRI compared with immediate CT was $7,483 per scaphoid fracture detected.	Immediate CT and MRI were found to be the most cost-effective strategies for diagnosing suspected scaphoid fractures.
Karl et al. (2015) US	**Study design:** Economic modelling using a decision analysis model (decision tree). **Follow-up:** No specific time horizon is defined.	**Interventions:** use of advanced imaging (CT or MRI) to initially diagnosis suspected scaphoid fractures (following an initial negative radiography). **Comparator:** Empiric cast immobilisation following a negative X-ray with orthopaedic follow-up and repeat X-ray two weeks post-injury.	Reference case was an employed patient who was 25 years of age with history and physical examination concerning for scaphoid fracture but with negative initial radiographs	**Costs:** Costs and benefits retrieved from literature. A societal perspective was used in the economic model. **QALYs:** Yes. **Economic evaluation:** Cost-effectiveness and cost-utility analyses.	**Cost-effectiveness/cost-utility:** Advanced imaging was dominant over empiric cast immobilisation. MRI was slightly more cost-effective than CT (ICER of $41,000/QALY and assuming a $100,000/QALY willingness to pay).	Advanced imaging was found to be dominant over the comparator. The decision to use either CT or MRI is a function of individual institutional costs and local test performance.
Tiel-van Buul et al. (1995)	**Study design:** Economic modelling using values retrieved from clinical	**Intervention(s)**: Three diagnostic strategies: 1) repeated X-ray up to week 2; 2) repeated X-ray up to week 6;	A total of 160 patients was included.	**Costs:** direct medical costs associated with the diagnostic (e.g. radiography, bone scintigraphy) and therapeutic	**Costs/Cost-effectiveness:** The mean cost per patient was $238, $276 and $275 for strategy 1, 2 and 3, respectively. The	The combined use of radiography and bone

The Netherlands	literature for: 1) the prevalence of scaphoid fractures among patients with suspected scaphoid fracture; 2) accuracy for different imaging modalities; and 3) percentage of non-union episodes given the type of therapy. **Follow-up:** Minimum follow-up period of one year.	and 3) X-ray, followed by bone scintigraphy (performed at least 3 days after the injury) in patients with negative findings in the X-ray. **Comparator**: imaginary scenario, assuming an initial radiography. If positive, patients are put in a plaster cast for 12 weeks. If the radiography is negative, all patients underwent bone scintigraphy and subsequent treatment is based on its findings.		(e.g. plaster cast, surgery for non-union) pathway were included. **Clinical outcomes**: mean period of immobilisation (measured in days) and non-union rate (measured as % of all patients). **QALYs:** None. **Economic evaluation:** Cost and cost-effective analyses, using the increment costs to save one non-union as the measure of cost-effectiveness.	incremental cost incurred to save one non-union by using bone scintigraphy is one-third of the price of repeated radiography at 6 weeks, hence the intervention is cost-effective (ICER not presented). **Other results:** Assuming a 44% prevalence rate, the mean immobilisation times were 5.7, 8.6 and 6.9 for strategy 1, 2 and 3, respectively. A non-union rate of 4.7%, 4.2% and 3.1% were assumed for strategy 1, 2 and 3, respectively.	scintigraphy is cost-effective.
Cost analyses (using economic models)						
Burns et al. (2013) UK	**Study design:** Descriptive analysis of retrospective data (clinical audit) + economic modelling for hypothetical intervention. **Follow-up:** Not specified.	**Intervention:** hypothetical early use of MRI. **Comparator:** clinical and radiographic diagnosis and patient's immobilisation in acute setting with subsequent periodic clinical and radiographic follow-up.	537 patients aged 13 years or older referred to fracture clinic at the Royal Infirmary of Edinburgh with a scaphoid-related injury.	**Costs:** Cost per type of injury (3 groups). Clinical notes were examined retrospectively and three injury groups were defined: true fractures, occult fractures, and suspected scaphoid injuries. **QALYs:** None **Economic evaluation:** not applicable (cost analysis).	**Costs:** There were 87 true fractures, 43 occult fractures, and 407 suspected injuries, incurring average treatment costs of £1,173 ($1,675), £773 ($1,104), and £384 ($548) respectively. The costs involved in the treatment of suspected scaphoid injuries were found to be higher than the additional cost of performing an MRI scan.	The introduction of early Magnetic Resonance Imaging (MRI) protocol would lead to an earlier definitive diagnosis and potentially be more cost-effective.
Ganeshali-ngam et al. (2013) Australia	**Study design:** Retrospective clinical audit of patients with suspected scaphoid fracture. A cost analysis is performed. **Follow-up:** Not specified.	**Intervention:** Patients with negative findings in the initial radiography underwent an MRI performed a mean of 6 days from the injury (range 1-21 days). Subsequent treatment is dependent on the MRI findings. **Comparator:** Traditional treatment algorithm, with immobilisation with removable scaphoid cast and periodic clinical follow-up 2.5 weeks after injury.	110 patients with suspected scaphoid fracture between August 2004 and March 2007.	**Costs:** Total cost savings associated with the use of MRI. Medical direct costs (MRI, radiography and bone scan) are considered, as well as societal costs due to time off work. **Clinical outcomes:** Percentage of patients that had an alteration to their management care as a result of the early MRI. **QALYs:** None. **Economic evaluation:** not applicable (cost analysis).	**Costs:** from a healthcare perspective, the use of MRI leads to an increase of $242 in costs. However, if loss earnings due to unnecessary immobilisation are included, it is estimated that the use of MRI will lead to cost saving of $1,655 per patient. **Other results:** 76% of patients with normal initial X-ray had an alteration to their management based on the MRI findings.	MRI in the assessment of suspected scaphoid fractures is found to change condition's clinical management and to be cost-effective.

Jenkins et al. (2008) UK	**Study design:** Economic modelling using the local prevalence of the condition and accuracy values for different imaging modalities retrieved from literature. **Follow-up:** No specific follow-up time is defined as patients are followed until a fracture is demonstrated by plain radiography or second line imaging or full resolution of the symptoms with no radiographic evidence of fracture is achieved.	**Interventions:** use of advanced imaging (CT, MRI, Bone Scintigraphy and Ultrasound) as the second line investigation for patients with suspected scaphoid fracture. **Comparator:** Empiric cast immobilisation following a negative radiography with orthopaedic follow-up and repeat radiography at 10-14 days post-injury.	The prevalence of scaphoid fractures was calculated using 200 consecutive patients with suspected scaphoid fracture.	**Costs:** Total cost of diagnosis and treatment of a patient with suspected scaphoid fracture. Costs such as staff, materials and overheads were considered and retrieved from local hospital (in Scotland) and the Scottish Health Statistics Cost book 2006. The total cost per patient was estimated based on: 1) the local prevalence; and 2) the accuracy levels of MRI, CT, Bone Scintigraphy and Ultrasound in the diagnosis of scaphoid fractures. **Clinical outcome:** True prevalence of scaphoid fractures among patients with suspected scaphoid fracture. **QALYs:** None **Economic evaluation:** not applicable (cost analysis).	**Costs:** The mean cost per patient was £302 ($482), £243 ($388), £113 ($180) and £202 ($322) for MRI, Bone Scintigraphy, Ultrasound and CT, respectively. These figures compare against the actual mean cost per patient using the standard management pathway is £204 ($326) for patients with confirmed scaphoid fracture and £132 ($211) for patients with no confirmed scaphoid fracture. **Other results:** The true prevalence of patients with scaphoid fracture amongst patients with suspected scaphoid fracture was 16%.	Clinicians should consider the use of alternative imaging modalities based on their local facilities and patient demographics.
Saxena at al. (2003) UK	**Study design:** Prospective clinical audit and economic modelling. **Follow-up:** Not specified, but includes follow-up until the plaster was removed and fractures ruled out in the last assessment.	**Intervention(s)**: Impact of 5 interventions is hypothesised: i) MRI scan on day 1; ii) Bone scan within a few days; iii) MRI scan within a few days; iv) MRI scan within a few days followed by a review of results on the same day; and v) MRI scan after a clinical and/or radiological examination in 2 weeks' time. **Comparator:** immobilisation with scaphoid plaster and periodic clinical and radiographic follow-up.	Clinical audit carried out in the Leighton General Hospital. A total of 85 eligible patients was considered.	**Costs:** Costs to the hospital budget in ruling out or diagnosing scaphoid fractures. Non-hospital costs, e.g. indirect costs associated with days off work were not considered. The resource use for the five alternative interventions is not based on data collected during the audit. **QALYs:** None. **Economic evaluation:** Cost-minimisation evaluation (cost analysis).	**Costs:** Projected costs for many of the alternative protocols involving routine use of scanning procedures were found to be lower or comparable to the comparator, immobilisation with periodic clinical and radiographic follow-up.	Protocols based on the use of advanced imaging (either MRI or bone scan) for suspected scaphoid fractures present comparable or lower overall costs compared to a protocol without advanced imaging.
Dorsay et al. (2001)	**Study design:** Economic modelling	**Intervention:** Screening MRI at the time of presentation in	Four actual case scenarios are	**Costs:** Direct cost associated with the intervention and the	**Costs:** Cost analysis suggests the two protocols are nearly	The cost analysis

US	using a decision analysis model (decision tree). **Follow-up:** Not specified but includes follow-up consultations.	patients with negative radiographic findings in the initial conventional radiography. **Comparator:** Immobilisation with scaphoid plaster and periodic clinical and radiographic follow-up.	presented and the impact of the intervention against the comparator (i.e. standard care) is hypothesised.	comparator. The utilisation of healthcare resources is hypothesised for four actual case scenarios. Cost data were based on charges from the US healthcare system. **QALYs:** None. **Economic evaluation:** not applicable (cost analysis only).	equivalent from a financial standpoint.	suggests that the two protocols are nearly equivalent from a financial point of view. Significant societal costs associated with productivity losses may be substantial.

Appendix IV. Patient experience questionnaire assessing the pre-test, the test and the post-test stages for both groups in the colon study.

		Very negative		Fairly negative		Neither positive / negative		Fairly positive		Very positive	
		N	%	N	%	N	%	N	%	N	%
About the bowel preparation (pre-test)											
Collecting / receiving bowel prep kit	OC (n= 69)	1	1.4%	4	5.8%	10	15%	14	20%	40	58%
	CTC (n= 45)	2	4.4%	1	2.2%	6	13%	12	27%	24	53%
Quality of the written instructions	OC (n= 69)	1	1.4%	2	2.9%	6	8.7%	21	30%	39	57%
	CTC (n= 44)	0	0.0%	0	0.0%	2	4.5%	10	23%	32	73%
Administering the bowel preparation yourself	OC (n= 69)	0	0.0%	1	1.4%	7	10%	23	33%	38	55%
	CTC (n= 45)	0	0.0%	1	2.2%	2	4.4%	12	27%	30	67%
How you found the process overall	OC (n= 68)	0	0.0%	6	8.8%	12	18%	27	40%	23	34%
	CTC (n= 45)	0	0.0%	0	0.0%	7	16%	16	36%	22	49%

		Still not fully recovered		More than 12 hours		Up to 12 hours		Up to 6 hours		Up to 3 hours		Up to 1 hour		Immediately	
		N	%	N	%	N	%	N	%	N	%	N	%	N	%
About the bowel test															
How long did it take to feel fully recovered?	OC (n= 70)	9	13%	10	14%	3	4.3%	7	10%	13	19%	15	21%	13	19%
	CTC (n= 45)	2	4.4%	1	2.2%	6	13%	3	6.7%	12	27%	10	22%	11	24%

		Not at all		A little		Moderately		Quite a bit		Extremely	
		N	%	N	%	N	%	N	%	N	%
Straightforward	OC (n= 69)	2	2.9%	4	5.8%	10	15%	19	28%	34	49%
	CTC (n= 42)	1	2.4%	1	2.4%	4	9.5%	17	41%	19	45%
Dignified	OC (n= 68)	2	2.9%	3	4.4%	14	21%	31	46%	18	27%
	CTC (n=38)	0	0.0%	3	7.9%	10	26%	11	29%	14	37%
Uncomfortable	OC (n= 66)	7	11%	26	39%	16	24%	8	12%	9	14%
	CTC (n= 38)	7	18%	10	26%	10	26%	6	16%	5	13%
Intrusive	OC (n= 65)	10	15%	24	37%	16	25%	9	14%	6	9.2%
	CTC (n= 38)	10	26%	16	42%	7	18%	3	7.9%	2	5.3%
Bearable	OC (n=66)	3	4.5%	2	3.0%	19	29%	26	39%	16	24%
	CTC (n=38)	2	5.3%	5	13%	11	29%	10	26%	10	26%

		A worse experience		As expected		A better experience	
		N	%	N	%	N	%
Overall experience about the diagnostic test							
How did you find the experience in comparison to your expectation?	OC (n= 70)	9	13%	31	44%	30	43%
	CTC (n= 45)	5	11%	23	51%	17	38%

		Very slightly or not at all		A little		Moderately		Quite a bit		Extremely	
		N	%	N	%	N	%	N	%	N	%
Feelings on the morning after the bowel test (post-test)											
Calm	OC (n= 67)	1	1.5%	3	4.5%	16	24%	29	43%	18	27%
	CTC (n= 42)	3	7.1%	2	4.8%	13	31%	14	33%	10	24%
Confident	OC (n= 66)	1	1.5%	5	7.6%	16	24%	31	47%	13	20%
	CTC (n= 40)	3	7.5%	2	5.0%	12	30%	16	40%	7	18 %
Safe	OC (n= 67)	1	1.5%	3	4.5%	12	18%	22	33%	29	43%
	CTC (n= 38)	1	2.6%	2	5.3%	7	18%	16	42%	12	32%
Relieved	OC (n= 66)	6	9.1%	4	6.1%	14	21%	14	21%	28	42%
	CTC (n= 42)	3	7.1%	3	7.1%	12	29%	14	33%	10	24%
Concerned	OC (n=65)	26	40%	19	29%	9	14%	7	11%	4	6.2%
	CTC (n=38)	9	24%	17	45%	6	16%	2	5.3%	4	11%
Anxious	OC (n=67)	27	40%	25	37%	9	13%	3	4.5%	3	4.5%
	CTC (n=41)	15	37%	17	42%	4	9.8%	3	7.3%	2	4.9%
Upset	OC (n=66)	52	79%	4	6.1%	7	11%	1	1.5%	2	3.0%
	CTC (n=38)	30	79%	4	11%	2	5.3%	1	2.6%	1	2.6%
Uncertain	OC (n=64)	29	45%	20	31%	8	13%	4	6.3%	3	4.7%
	CTC (n=40)	18	45%	9	23%	9	23%	1	2.5%	3	7.5%

Appendix V. Strengths and weaknesses of different types of studies used in the economic evaluations.

Randomised controlled trials (RCTs)

RCTs are commonly considered as the gold standard to establish the efficacy of a given technology. This is due to its implicit design that eliminates different types of biases, particularly selection bias. It has however been argued that RCTs might not be the best type of study on which to base an economic evaluation (Sculpher et al. 2006). RCTs' high internal validity due to its randomised and controlled design can limit external validity and generalisability to other healthcare groups, providers or healthcare settings.

Strengths:

- Unbiased estimates of treatment effects. The RCT design aims to ensure that differences in outcomes are attributed to the intervention being evaluated. This is achieved using a randomised allocation of participants to the intervention or the control group (e.g. the placebo group in drug trials), a well-defined participant inclusion and exclusion criteria and tight protocol-driven processes.
- Prospective collection of both clinical outcomes and resource use data at the patient-level. Access to individual data allows for the utilisation of different statistical and economic techniques to explore the relationship between costs and outcomes such as health-related quality of life (Petrou and Gray 2011; Petrou 2012; Glick et al. 2007).
- Adequately powered and scientifically rigorous.

Weaknesses:

- Single-centre RCTs might not provide all the information required and may not be generalisable (Woolacott et al. 2017). RCTs are generally designed to assess the technology's efficacy and therefore relevant data from the economic perspective might be missing. Sculpher et al. (2006), based on two cardiology studies, noted that RCTs lacked patient reported outcomes such as quality of life (e.g. QALYs) and failed to capture all relevant resource use data. This might lead to the need to use evidence synthesis and modelling techniques to incorporate these costs and effects into the economic evaluation.
- Controlled setting and inclusion criteria. The trial's inclusion criteria might not be representative of the heterogeneous population that might benefit from the intervention in an NHS-based environment, thus potentially rendering the decision inappropriate.

- Insufficient time horizon. Most trial-based economic evaluations present a follow-up period shorter than the time horizon considered in an economic model. This is particularly the case for RCTs in chronic diseases that might require valuation of long-term or lifetime costs and effects (Sculpher et al. 2006; O'Sullivan, Thompson, and Drummond 2005). This limitation means that cost and effects after the trial's follow-up period are typically derived from decision model extrapolation from cost and outcome observed in the trial. Sculpher et al. (2006) pointed out that the latter might not be a problem if the within-trial incremental analysis already satisfies the decision makers' willingness to pay threshold.
- Artificially enhanced patient compliance rates. The RCT setting might lead to patient compliances rates with the clinical management that are far superior to what might be seen in a real-world clinical practice (S. D. Ramsey et al. 2015).
- Might not provide information on final endpoints. RCTs might be based on intermediate clinical endpoints rather than the actual final outcome required for the economic evaluation. This is quite often the case in the medical imaging field, where intermediate outcomes (e.g. sensitivity and specificity) are often reported in RCTs rather than actual clinical outcomes (e.g. mortality rates).
- Might not compare all relevant alternatives (Sculpher et al. 2006; O'Sullivan, Thompson, and Drummond 2005). Given the research costs and feasibility issues, RCTs typically compare a new technology against standard care. However, other technologies, often used to a lesser degree, are not included. Thus, RCTs typically do not compare all existing alternatives. In this context, an economic evaluation based on a single RCT produces a partial analysis (Sculpher et al. 2006; Cook 2015).
- Lack of relevance to the decision maker. Quite often RCTs are conducted by the technology's manufacturer, that might target important healthcare markets or markets and providers where recruitment is quicker and/or at lower costs (Sculpher et al. 2006; O'Sullivan, Thompson, and Drummond 2005). However, an RCT from a non-UK background might present considerable differences in patient case-mix and routine clinical practice, posing generalisability concerns.

Real-world evidence (RWE)

More recently, several authors have argued in favour of pragmatic trials-based economic evaluations (Buxton et al. 1997; O'Sullivan, Thompson, and Drummond 2005; Glick et al. 2007; Marshall and Hux 2009; Calvert, Wood, and Freemantle 2011; Petrou 2012; S. D. Ramsey et al. 2015). This opportunity derives from the naturalistic study design of pragmatic trials, explicitly aimed at: (i) assessing the intervention's effectiveness in real-world clinical practice rather than its efficacy; and (ii) increasing the generalisability of the findings, with a population and clinical practice

representative of a real-world setting. In the context of health economics, RWE has also been described as data used to support an economic evaluation based on studies that were not RCTs (Garrison et al. 2007). Real-world studies ranged from: large simple trials, registries, administrative data to electronic health records reviews (Garrison et al. 2007).

Strengths:

- Large simple trials are, in effect, pragmatic RCTs with broad inclusion criteria, more reflective of clinical practice. Similar to conventional RCTs, pragmatic trials have the strengths associated with the randomised allocation of participants. Their pragmatic design however means that a wider population is recruited, allowing for a better understanding of resource use for a more heterogeneous population. In addition, the use of pragmatic trials has been associated with the inclusion of an economic component in the trial design (Calvert, Wood, and Freemantle, 2011).
- Registries are prospective, non-randomised, observational cohort studies (Garrison et al. 2007) that involve data collection of uniform clinical, economic and/or patient-reported outcomes. Compared to pragmatic trials, registries tend to be cheaper and include a large and more heterogeneous population and longer periods of follow-up. Hence, registries are more likely to capture final outcomes and not just intermediate outcomes.
- Observational case-control studies are retrospective studies. Observational studies can be very useful complementing information obtained from RCTs, allowing researchers to analyse and generalise to subpopulations not included in the RCTs inclusion criteria (Silverman, 2009). However, given the non-randomised design of observational studies, appropriate statistical methods (e.g. regression and matching methods) should be used to address potential selection biases. Furthermore, even if the right statistical methods are used, decision makers need to be cautious about making inferences based on results from observational studies (Freemantle et al. 2013; Woolacott et al. 2017).
- Administrative data considers the retrospective, or, if feasible, real-time data collection, of data typically used for reimbursement purposes (Garrison et al. 2007). This information not only captures resource use but also diagnoses and procedure codes with information on participants (e.g. comorbidities). Administrative data can provide useful retrospective cross-sectional (at a given point in time) or longitudinal (over a specific follow-up period) analyses of clinical and economic outcomes for a given patient, group or population (Garrison et al. 2007). Given its low cost and ease to perform, retrospective analysis of administrative data might provide useful insights into plausible associations between any given intervention and clinical and economic outcomes. If participants grant consent, the use of administrative data can be combined with data from registries.

- Electronic health records are important sources of real-world data as contain disease-specific information at a patient or group level and are accessible, standardise reliable sources of data (Franklin and Thorn 2019). This type of data [e.g. Hospital Episode Statistics (HES)] is commonly used alongside other types of real-world data.

Weaknesses:

- Large simple trials (pragmatic RCTs) present typically larger sample sizes, being more time and cost consuming to perform (Garrison et al. 2007). Consistent with conventional RCTs, pragmatic trials do not eliminate the remaining weaknesses associated with RCTs, e.g. not comparing all alternatives or inappropriate time horizon. The choice of outcomes (primary and secondary outcomes) and respective measurement methods might be constrained in favour of the trial's feasibility (Welsing et al. 2017).
- Registries do not include random allocation of participants to an intervention. Thus, results need to be interpreted with caution due to the intrinsic limitations of observational studies (Garrison et al. 2007) and the use of analytic methods to address imbalance in baseline characteristics should be considered.
- Retrospective administrative data analyses rely on data entries outside the scope of the researcher that are primarily developed for reimbursement purposes. Any economic evaluation based on administrative data could lead to biased findings if its data quality is poor. In fact, administrative data quality is usually questioned due: missing data, particularly data not missing at random; coding errors or potential data 'gaming'; the utilisation of charges as a proxy for healthcare costs; the variability between healthcare providers and its coding practices; the limited availability of clinical, socioeconomic and demographic characteristics; and lack of data on health outcomes and symptoms (Garrison et al. 2007). Lastly, even when a research question is clearly articulated, it might be hard to identify the entire population at risk or likely to benefit from the intervention (i.e. to determine the population denominator).
- Electronic health records typically lack data integration to allow the estimate of healthcare utilisation on a patient-level basis. This usually requires data agreements to be put in place and time-consuming data extraction periods (Franklin and Thorn 2019)

Decision models

Decision models are based on mathematical approaches to simulate uncertain real world scenarios and support the decision making. Buxton et al. (1997) proposed two scenarios where modelling is appropriate. First, when in the presence of an innovative technology or when few data are available, decision models are used as a first resort for hypothesis generation and inform the design of future studies (Buxton et al. 1997). Second, modelling should be used at the opposite end of the spectrum, as a last resort when RCTs studies are not feasible or, when possible, do not provide required information. Decision modelling have sometimes been wrongly described as an alternative rather than a complement to experimental and real-world studies (Sculpher et al. 2006).

Strengths:

- Unlike RCTs or pragmatic RCTs that typically present a head-to-head comparison (e.g. new intervention vs placebo or current standard care), decision models allow for the inclusion of all relevant alternatives associated with the decision problem.
- Extrapolate beyond observed data. Given that clinical trials generally provide comparative data for a limited period of follow-up, economic models can extrapolate data (e.g. mortality rates or disease progression) beyond the observed data (Buxton et al. 1997).
- Link intermediate with final outcomes. Depending on the clinical condition, it is not uncommon for observed data to evaluate an intermediate endpoint rather than final endpoints. One example pointed out by Buxton et al. (1997) was the use of variations in total serum cholesterol (intermediate outcome) as a surrogate for the assessment of hypercholesterolaemia (final outcome). Economic decision models are then used to link intermediate with final outcomes. However, surrogate endpoints tend to over-estimate treatment effects (Woolacott et al. 2017).
- Increased generalisability of results. Economic evaluations are performed to reflect the real-world setting encountered by the decision maker. This means that is usually necessary to generalise data, from trials (efficacy) to real-world clinical practice (effectiveness) and also from different healthcare providers (hospital A and hospital B) or even different healthcare systems (e.g. US, UK) (Buxton et al. 1997; Woolacott et al. 2017).
- Data synthesis and multiple comparison. Economic models are used to synthesise data from multiple sources (e.g. different RCTs, systematic literature reviews or observational studies) (Garattini et al. 2016). In the context of multiple RCTs and RWE, decision modelling based on data synthesis (e.g. meta-analyses) allows the comparison of all alternatives and even a head-to-head comparisons between interventions where

no trial data exists (Buxton et al. 1997; Sculpher et al. 2006) or the use of data from studies with variable follow-up periods (Sculpher et al. 2006).

- Uncertainty analysis. Economic models can be useful supporting the decision maker process even in the absence of observed data (Buxton et al. 1997; Woolacott et al. 2017). Furthermore, the use of deterministic and Bayesian probabilistic sensitivity analyses across multiple model parameters enables the decision make to evaluate under which scenarios the intervention should be adopted (Ades et al. 2006).

Weaknesses:

- Inadequate clinical data. Economic modelling based on clinical data is intrinsically associated with the quality of the respective clinical data and its study design. Economic models based on flawed or biased clinical data are likely to provide a biased decision to decision makers.
- Inadequate use of statistical methods associated with data derived from multiple studies (heterogeneity) (Ades et al. 2006).
- Overall uncertainty associated with its findings, requiring careful interpretation by decision makers (Woolacott et al. 2017).

Appendix VI. Key literature results, detailing the incidence of suspected scaphoid fracture and the sensitivity and specificity for different imaging modalities.

Reference	Incidence	Sensitivity	Specificity	Notes
Yin et al., 2010. "Diagnosis suspected scaphoid fractures: a systematic review and meta-analysis."	Variable: From 5% up to 50%.	MRI: 96% (95% CI: 91%-99%) CT: 93% (95% CI: 83%-98%)	MRI: 99% (95% CI: 96%-100%) CT: 99% (95% CI: 96%-100%)	26 studies were included in the systematic review and meta-analysis. Variable average age of patients: from 23 up to 44 years old.
Ring and Lozano-Calderón, 2008. "Imaging for suspected scaphoid fracture."	11% of wrist injuries are suspected scaphoid fractures. Out of these, 7% are true scaphoid fractures.	MRI: 98% (No CI identified) CT: 94% (No CI identified)	MRI: 99% (No CI identified) CT: 96% (No CI identified)	Values proposed based on 22 and 8 original research publications to address the performance of MRI and CT, respectively.
Memarsadeghi et al., 2006."Occult scaphoid fractures: comparison of multidetector CT and MR Imaging – Initial Experience"	38% (11 fractures out of 29 patients)	MRI: 100% (No CI identified) CT: 73% (No CI identified)	MRI: 100% (No CI identified) CT: 100% (No CI identified)	Compared to MRI, multidetector CT performs better in the detection of cortical scaphoid fractures but is worst in depicting solely trabecular injuries.
Hackney and Dodds 2011". Assessment of scaphoid fracture healing".	N/A	Follow-up radiography: 9 to 49%. CT: 72%. No CI identified)	CT: 80%. No CI identified)	Review article. Not a case study.
Mallee et al., 2011."Comparison of CT and MRI for diagnosis of	Average from meta-analyses of patients with normal X-ray:	CT: 67% No CI identified)	CT: 96% No CI identified) MRI: 89% No CI identified)	Case study with a small sample (n=34). It is also relevant to point out that a 1.0T open field MRI was used (leading to potential motion

suspected scaphoid fractures".	16%; Case study: 18%	MRI: 67% No CI identified)		artefacts and lower image quality).
Brydie and Raby, 2003. "Early MRI in the management of clinical scaphoid fracture".	Average from meta-analyses of patients with normal X-ray: 16% Case study: 19%	N/A	N/A	Average patient age was 36 years. A 0.2T dedicated extremity scanner was used and the images acquired within 14 days of injury. The management of 92% of the patients was altered due to the use of MRI. 106 patients without any positive MRI finding were discharged. 15 patients were discharged after one clinic visit.
Low and Raby, 2005. "Can follow-up radiography for acute scaphoid fracture still be considered a valid investigation?"	N/A	Follow-up radiography: 9-49%.	Follow-up radiography: 80-93%.	Follow-up radiography *should not* be considered a valid investigation for the detection of scaphoid fracture.
Pillai and Jain (2005). "Management of clinical fractures of the scaphoid: results of an audit and literature review."	6.6% of scaphoid fractures (5/75).	N/A	N/A	Method: prospective study with 90 patients with clinical signs suggestive of scaphoid injury.
Raby (2001). "Magnetic resonance imaging of suspected scaphoid fractures using a low field dedicated extremity MR system."	13.2% of scaphoid fractures (7/53)	N/A	N/A	0.2T extremity MR system was used. Management was altered in 69%.

Appendix VII. Description of the five domains and respective constructs considered in the CRIF framework applied to the three implementation projects considered in this thesis (suspected scaphoid fracture, chronic headache and colorectal cancer pathways).

Domain	Constructs	Suspected scaphoid fracture	Chronic headache	Colorectal cancer (CRC)
Intervention characteristics	Intervention source	Internally developed by orthopaedics and radiologists.	Internally developed by neurologists and radiologists.	Internally developed by colorectal clinicians and radiologists.
	Evidence strength and quality	Based on a pragmatic randomised clinical trial, perceived as high-quality evidence.	Based on published data and local observational, prospective studies. The research studies were perceived as high-quality and allowed a better understanding of the two existing clinical pathways for both clinicians and non-clinicians alike.	
	Relative advantage	MRI perceived as the best diagnostic test to diagnose suspected scaphoid fractures.	Direct access to MRI perceived as a valid alternative to some patients.	Mixed perception among clinicians: CTC perceived as inferior or equal to reference colonoscopy.
	Adaptability	Provision of MRI was not continuous (24/7). Local plans were made to provide equal care to all patients regardless of time of presentation to ED.	Both clinical pathways already constituted standard care. However, changes were made to promote uptake of the intervention, both in the chronic headache pathway (new MRI reporting templates) and the CRC pathway (new bowel preparation workflows).	
	Trialability	Local pilot conducted prior to roll-out of the new pathway.	Not applicable. Both clinical pathways already constituted standard care.	
	Complexity	Disruptive intervention, perceived as complex, and involving multiple stakeholders	Implementation not disruptive in nature. Complexity of the chronic headache implementation deemed high given the multiple stakeholders at primary and secondary care.	

	Design quality and packaging:	Similar across all implementation projects. Clinical and cost findings from the real-world studies were presented in detail to the different clinical and managerial teams.		
	Cost:	The intervention was associated with lower NHS costs. Implementation costs included time taken off by clinicians and non-clinicians to adopt the new pathway.	The intervention was associated with lower NHS costs. Implementation costs were higher compared to other projects a GP training programme was considered.	The intervention was associated with lower NHS costs. Implementation costs were marginal, associated with the set-up of a new workflow for dispensing the bowel preparation.
Outer settings	Patient needs and resources	The implementation aimed to improve patient satisfaction by reducing the time from presentation to diagnosis and avoiding unnecessary wrist immobilisation.	The implementation aimed to improve uptake with the direct access to brain MRI and decrease patient waiting times (improve access to care).	The change in the bowel preparation workflow was driven by patient needs.
	Cosmopolita-nism:	Guy's and St Thomas' NHS Foundation Trust (GSTT) is a central hospital with different network links to multiple bespoke hospitals. Multiple clinicians at GSTT were responsible to oversee the integration of care between different healthcare providers and have been exposed to different implementation initiatives.		
	Peer pressure	Given the non-competitive nature of the UK healthcare market and the innovative nature of the interventions (i.e. not standard care across the NHS), there was no significant peer pressure associated with the implementation initiatives.		
	External policies and incentives	Potential increase in the 4-hour ED target and no unbundled tariff for MRI scans performed as part of the ED presentation represented two external negative financial incentives.	Clinical guidelines promoted the utilisation of GP direct access to imaging. Existing unbundled MRI tariff provided external positive financial incentives.	Clinical guidelines did not promote the equivalence of CTC and OC (negative external policy). However, waiting times associated with OC promoted the use of alternatives to

				avoid financial penalties (negative incentives).
Inner settings	Structural characteristics	The implementation of novel initiatives in the ED was challenging due to: a high turnover of human resources (teams are less stable); large teams; and considerable internal pressure to deliver.	The Radiology department, main responsible for the delivery of direct access to MRI (headache pathway) and CTC (CRC pathway), although large in size compared to other Hospitals, was characterised for decentralising decision-making to local, with smaller groups directly responsible for delivery of care.	
	Network and communica-tions	Existing internal networks, relationships and communication channels between the different directorates (ED, Radiology and Orthopaedics) were suboptimal.	Existing external communication channels with GPs were better developed in the Neurology and Gastroenterology departments compared to Radiology. Lack of appropriate communication channels for interaction between GPs (referrer) and the MRI/CTC services. Existing internal networks and relationships among directorates were based on informal relationships between key clinicians.	
	Culture:	GSTT has recently embarked in a journey of cultural change where adaptive and continuous change is perceived as essential to ensure the organisation's sustainability. However, compared to other healthcare providers (particularly in the USA), GSTT is still in the early stages of cultural development.		
	Implementation climate	The existing clinical pathway was perceived as of poor quality (tension for change). The implementation initiative was aligned with the individuals and organisation's values and promoted sense of belonging among the teams.	The two implementation initiatives aimed to increase uptake of an existing pathway (i.e. no particular tension for change). In both cases, the initiatives were compatible with the individuals and organisation's values and teams were given the latitude to improve care for patients. The suspected CRC pathway was considered a priority for the organisation (cancer care).	

		However, did not constitute a priority to the organisation.	
	Readiness for implementation	Given its innovative nature, this implementation initiative required senior leader engagement (e.g. GSTT's CEO and Medical Director attended multiple ED huddles at 8am). Furthermore, given its complexity, appropriate resources (clinician and non-clinicians' time) were dedicated.	The level of senior engagement required for the implementation was considered to be desirable (not essential as with the scaphoid initiative). To maximise the implementation's success, individuals and groups of individuals were given resources and information required to improve clinical practice.
Characteristics of the individuals involved	Knowledge and beliefs about the intervention:	Individuals, as well as organisations, present a natural resistance to change. The latter may be particularly true in clinical areas where disruptive innovation is less common. Individuals in the Radiology Department are commonly exposed to new technologies and were expected to be less resistant to change.	
	Self-efficacy:	Individuals at GSTT were believed to be quite self-confident of their own abilities to implement change across their models of care.	
	Individual stage of change:	Individuals (clinicians and non-clinicians) involved in the empirical studies were towards the last stages of the change model (action and maintenance stages). However, other members of staff were at very different stages, with some, particular in the ED, believed to be resistant to change (pre-contemplation and contemplation stages).	
	Individual identification with organisation:	Staff satisfaction levels at GSTT are above the UK average (measured by national staff surveys). Moreover, given the hospital's teaching role and close links to academia (KCL), moderate to high levels of organisation commitment were anticipated.	

	Other personal attributes:	GSTT, being one the main hospitals in the UK and also based in central London, presents multi-cultural staff members and patients. Individuals were anticipated to be motivated, tolerant and with intellectual ability.	
Process of implementation	Planning:	Detailed formal implementation plans were considered for each initiative. These were based in A3 project sheets and disseminated across the different individuals and departments involved.	
	Engaging:	Volunteered clinical leaders within each directorate were pivotal at cascading down key information and actions of the implementation plan. This was achieved by the combination of: internal widespread of research findings from the research studies; informal and formal operational meetings; multiple education and drop-in sessions; and workshops with multiple stakeholders (from administrators to clinicians)	
	Executing:	Implementation plans were, to a large degree, followed by the different teams. However, there were some delays in some key activities, particularly: the definitive roll-out of the scaphoid pathway (postponed twice due to discrepancies on who is responsible to communicate the MRI findings under circumstances); and the revision of the 'incidental findings' section in the brain MRI report (ongoing process).	
	Reflecting and evaluating:	Monthly meetings for 12 months following the implementation roll-out and education and drop-in sessions for the first 2 months. These meetings are based on quantitative data (e.g. number of MRI referrals, proportion of patients undergoing MRI, distribution of clinical findings, cost savings to the NHS). Ongoing discussions with CCGs also aim at revising existing reimbursement tariffs associated with the intervention.	Key opinion leaders involved in the implementation project were given access to a purpose built dashboard where they could see key implementation data such as: absolute number of referral to either brain MRI or CTC; number of days elapsed between referrals and respective tests; number of days from test to report. These aimed at providing these members of staff with enough evidence to continuously monitor the progress of the implementation project and ensure the intervention is sustained over time.

References

Ades, A. E., Mark Sculpher, Alex Sutton, Keith Abrams, Nicola Cooper, Nicky Welton, and Guobing Lu. 2006. “Bayesian Methods for Evidence Synthesis in Cost-Effectiveness Analysis.” *PharmacoEconomics* 24 (1): 1–19. https://doi.org/10.2165/00019053-200624010-00001.

Agus, A. M., P. McKavanagh, L. Lusk, R. M. Verghis, G. M. Walls, P. A. Ball, T. R. Trinick, M. T. Harbinson, and P. M. Donnelly. 2016. “The Cost-Effectiveness of Cardiac Computed Tomography for Patients with Stable Chest Pain.” *Heart (British Cardiac Society)* 102 (5): 356–62. https://doi.org/10.1136/heartjnl-2015-308247.

Anonychuk, Andrea, Graham Beastall, Simon Shorter, Regina Kloss-Wolf, and Peter Neumann. 2012. “A Framework for Assessing the Value of Laboratory Diagnostics.” *Healthcare Management Forum* 25 (3_suppl): S4–11. https://doi.org/10.1016/j.hcmf.2012.07.015.

Apthorp, L.A., C.A. Daly, I.D. Morrison, and S. Field. 1998. “Direct Access MRI for General Practitioners — Influence on Patient Management.” *Clinical Radiology* 53 (1): 58–60. https://doi.org/10.1016/S0009-9260(98)80036-6.

Atkin, Wendy, Edward Dadswell, Kate Wooldrage, Ines Kralj-Hans, Christian von Wagner, Rob Edwards, Guiqing Yao, et al. 2013. “Computed Tomographic Colonography versus Colonoscopy for Investigation of Patients with Symptoms Suggestive of Colorectal Cancer (SIGGAR): A Multicentre Randomised Trial.” *The Lancet* 381 (9873): 1194–1202. https://doi.org/10.1016/S0140-6736(12)62186-2.

Bainbridge, Sara. 2018. “CAPACITY TO DIAGNOSE?” https://www.cancerresearchuk.org/sites/default/files/mar18_capacity_to_diagnose.pdf.

Baker, Laurence C., Scott W. Atlas, and Christopher C. Afendulis. 2008. “Expanded Use of Imaging Technology and the Challenge of Measuring Value.” *Health Affairs (Project Hope)* 27 (6): 1467–78. https://doi.org/10.1377/hlthaff.27.6.1467.

Balas, E. A., and S. A. Boren. 2000. “Managing Clinical Knowledge for Health Care Improvement.” *Yearbook of Medical Informatics*, no. 1: 65–70.

Barber, Julie A., and Simon G. Thompson. 2000. “Analysis of Cost Data in Randomized Trials: An Application of the Non-Parametric Bootstrap.” *Statistics in Medicine* 19 (23): 3219–36. https://doi.org/10.1002/1097-0258(20001215)19:23<3219::AID-SIM623>3.0.CO;2-P.

Bateman, David. 2011. “The Future of Neurology Services in the UK.” *Practical Neurology* 11 (3): 134–35. https://doi.org/10.1136/practneurol-2011-000009.

Bergh, Torbjørn Hiis, Tommy Lindau, Lars Atle Soldal, Soosaipillai V. Bernardshaw, Mehdi Behzadi, Knut Steen, and Christina Brudvik. 2013. “Clinical Scaphoid Score (CSS) to Identify Scaphoid Fracture with MRI in Patients with Normal x-Ray after a Wrist Trauma.” *Emergency Medicine Journal*, May, emermed-2012-202219. https://doi.org/10.1136/emermed-2012-202219.

Berrington de Gonzalez, A., Kwang Pyo Kim, and Judy Yee. 2010. “CT Colonography: Perforation Rates and Potential Radiation Risks.” *Gastrointestinal Endoscopy Clinics of North America* 20 (2): 279–91. https://doi.org/10.1016/j.giec.2010.02.003.

Black, William C., Ilana F. Gareen, Samir S. Soneji, JoRean D. Sicks, Emmett B. Keeler, Denise R. Aberle, Arash Naeim, et al. 2014. “Cost-Effectiveness of CT Screening in the National Lung Screening Trial.” *New England Journal of Medicine* 371 (19): 1793–1802. https://doi.org/10.1056/NEJMoa1312547.

Bloudek, L. M., M. Stokes, D. C. Buse, T. K. Wilcox, R. B. Lipton, P. J. Goadsby, S. F. Varon, et al. 2012. “Cost of Healthcare for Patients with Migraine in Five European Countries: Results from the International Burden of Migraine Study (IBMS).” *The Journal of Headache and Pain* 13 (5): 361–78. https://doi.org/10.1007/s10194-012-0460-7.

Boardman, H F, E Thomas, P R Croft, and D S Millson. 2003. “Epidemiology of Headache in an English District.” *Cephalalgia: An International Journal of Headache* 23 (2): 129–37.

Boer, Michiel R de, Wilma E Waterlander, Lothar DJ Kuijper, Ingrid HM Steenhuis, and Jos WR Twisk. 2015. “Testing for Baseline Differences in Randomized Controlled Trials: An Unhealthy Research Behavior That Is Hard to Eradicate.” *The International Journal of Behavioral Nutrition and Physical Activity* 12 (January). https://doi.org/10.1186/s12966-015-0162-z.

Bowel Cancer UK. 2019. “Early Diagnosis Campaign.” Bowel Cancer UK. 2019. https://www.bowelcanceruk.org.uk/campaigning/early-diagnosis/.

Briggs, Andrew, Karl Claxton, and Mark Sculpher. 2011. *Decision Modelling for Health Economic Evaluation*. Oxford University Press.

Brooks, S, F Cicuttini, S Lim, D Taylor, S Stuckey, and A Wluka. 2005. “Cost Effectiveness of Adding Magnetic Resonance Imaging to the Usual Management of Suspected Scaphoid Fractures.” *British Journal of Sports Medicine* 39 (2): 75–79. https://doi.org/10.1136/bjsm.2003.007435.

Brydie, A, and N Raby. 2003. “Early MRI in the Management of Clinical Scaphoid Fracture.” *The British Journal of Radiology* 76: 296–300.

Burns, M. J., S. A. Aitken, D. McRae, A. D. Duckworth, and A. Gray. 2013. “The Suspected Scaphoid Injury: Resource Implications in the Absence of Magnetic Resonance Imaging.” *Scottish Medical Journal* 58 (3): 143–48. https://doi.org/10.1177/0036933013496950.

Buxton, Martin J., Michael F. Drummond, Ben A. Van Hout, Richard L. Prince, Trevor A. Sheldon, Thomas Szucs, and Muriel Vray. 1997. “Modelling in Ecomomic Evaluation: An Unavoidable Fact of Life.” *Health Economics* 6 (May): 217–27. https://doi.org/10.1002/(SICI)1099-1050(199705)6:3<217::AID-HEC267>3.0.CO;2-W.

Cake, Russell, Peter Cavanagh, and Ben Gordon. 2015. “HORIZON SCANNING HORIZON SCANNING: An Evaluation of Imaging An Evaluation of Imaging Capacity across the NHS in England.” Cancer Research UK. https://www.cancerresearchuk.org/sites/default/files/horizon_scanning_-_final.pdf.

Callaghan, Brian C, Kevin A Kerber, Robert J Pace, Lesli Skolarus, Wade Cooper, and James F Burke. 2015. “Headache Neuroimaging: Routine Testing When Guidelines Recommend against Them.” *Cephalalgia* 35 (13): 1144–52. https://doi.org/10.1177/0333102415572918.

Callaghan, Brian C., Kevin A. Kerber, Robert J. Pace, Lesli E. Skolarus, and James F. Burke. 2014. “Headaches and Neuroimaging: High Utilization and Costs despite Guidelines.” *JAMA Internal Medicine* 174 (5): 819–21. https://doi.org/10.1001/jamainternmed.2014.173.

Calvert, Melanie, John Wood, and Nick Freemantle. 2011. “Designing ‘Real-World’ Trials to Meet the Needs of Health Policy Makers at Marketing Authorization.” *Journal of Clinical Epidemiology* 64 (7): 711–17. https://doi.org/10.1016/j.jclinepi.2010.12.010.

Cancer Research UK. 2015. “Bowel Cancer Incidence Statistics.” Cancer Research UK. May 15, 2015. https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/bowel-cancer/incidence.

Centre for Reviews and Dissemination. 2009. “Systematic Reviews: CRD’s Guidance for Undertaking Reviews in Health Care.” http://www.york.ac.uk/inst/crd/pdf/Systematic_Reviews.pdf.

Cepeda, M. S. 2003. “Comparison of Logistic Regression versus Propensity Score When the Number of Events Is Low and There Are Multiple Confounders.” *American Journal of Epidemiology* 158 (3): 280–87. https://doi.org/10.1093/aje/kwg115.

Clemens, Maria. 2017. “Technology And Rising Health Care Costs.” Forbes. 2017. https://www.forbes.com/sites/forbestechcouncil/2017/10/26/technology-and-rising-health-care-costs/.

Coleman, MP, D Forman, H Bryant, J Butler, B Rachet, C Maringe, U Nur, et al. 2011. “Cancer Survival in Australia, Canada, Denmark, Norway, Sweden, and the UK, 1995-2007 (the International Cancer Benchmarking Partnership): An Analysis of Population-Based Cancer Registry Data.” *Lancet* 377 (9760): 127–38. https://doi.org/10.1016/S0140-6736(10)62231-3.

Collie, D A, R J Sellar, J P Steyn, and R E Cull. 1999. “The Diagnostic Yield of Magnetic Resonance Imaging (MRI) of the Brain and Spine Requested by General Practitioners: Comparison with Hospital Clinicians.” *The British Journal of General Practice* 49 (444): 559–61.

Community Headache Service. 2011. “2011_Headache Business Case_public.” http://www.exeterheadacheclinic.org.uk/sitebuildercontent/sitebuilderfiles/lambeth_headache_business_case_public_copy_2011.pdf.

Consort. 2010. “Consort - Welcome to the CONSORT Website.” 2010. http://www.consort-statement.org/.

Cook, John R. 2015. “Merging Regulatory and Reimbursement Needs in Clinical Trials.” *Value in Health* 18 (2): 145–46. https://doi.org/10.1016/j.jval.2015.02.007.

Crabb, Nick. 2011. “The NICE Diagnostics Assessment Programme.” In . Oxford.

Curtis, L. 2012. “PSSRU: Unit Costs of Health & Social Care 2012.” Personal Social Services Research Unit.

Curtis, L. 2013. “Unit Costs of Health and Social Care 2013 (PSSRU).” http://www.pssru.ac.uk.

Curtis, L., and A. Burns. 2017. “Unit Costs of Health and Social Care 2016 (PSSRU).” https://www.pssru.ac.uk/project-pages/unit-costs/unit-costs-2016/.

Curtis, L, and A Burns. 2018. “Unit Costs of Health and Social Care 2018 (PSSRU).” http://www.pssru.ac.uk.

Damschroder, Laura J., David C. Aron, Rosalind E. Keith, Susan R. Kirsh, Jeffery A. Alexander, and Julie C. Lowery. 2009. “Fostering Implementation of Health Services Research Findings into Practice: A Consolidated Framework for Advancing Implementation Science.” *Implementation Science* 4 (1): 50. https://doi.org/10.1186/1748-5908-4-50.

Dekkers, Tanja, Aleksander Prejbisz, Leo J. Schultze Kool, Hans J. M. M. Groenewoud, Marieke Velema, Wilko Spiering, Sylwia Kołodziejczyk-Kruk, et al. 2016. “Adrenal Vein Sampling versus CT Scan to Determine Treatment in Primary Aldosteronism: An Outcome-Based Randomised Diagnostic Trial.” *The Lancet. Diabetes & Endocrinology* 4 (9): 739–46. https://doi.org/10.1016/S2213-8587(16)30100-0.

Department of Health. 2012a. “Direct Access to Diagnostic Tests for Cancer: Best Practice Referral Pathways for General Practitioners.” https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/216503/dh_133511.pdf.

———. 2012b. “Direct Access to Diagnostic Tests for Cancer: Best Practice Referral Pathways for General Practitioners.” https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/216503/dh_133511.pdf.

Department of Health and Social Care. 2016. “NHS Reference Costs.” GOV.UK. 2016. https://www.gov.uk/government/collections/nhs-reference-costs.

Dettori, Joseph. 2010. “The Random Allocation Process: Two Things You Need to Know.” *Evidence-Based Spine-Care Journal* 1 (3): 7–9. https://doi.org/10.1055/s-0030-1267062.

Devlin, Nancy J., Koonal K. Shah, Yan Feng, Brendan Mulhern, and Ben van Hout. 2018. “Valuing Health-Related Quality of Life: An EQ-5D-5L Value Set for England.” *Health Economics* 27 (1): 7–22. https://doi.org/10.1002/hec.3564.

Dorsay, T. A., N. M. Major, and C. A. Helms. 2001. “Cost-Effectiveness of Immediate MR Imaging versus Traditional Follow-up for Revealing Radiographically Occult Scaphoid Fractures.” *AJR. American Journal of Roentgenology* 177 (6): 1257–63. https://doi.org/10.2214/ajr.177.6.1771257.

Drummond, M. F. 1996. “The Future of Pharmacoeconomics: Bridging Science and Practice.” *Clinical Therapeutics* 18 (5): 969–78.

Drummond, Michael, Adrian Griffin, and Rosanna Tarricone. 2009. “Economic Evaluation for Devices and Drugs—Same or Different?” *Value in Health* 12 (4): 402–4. https://doi.org/10.1111/j.1524-4733.2008.00476_1.x.

Drummond, Michael, Mark Sculpher, George Torrance, Bernie O'Brien, and Greg Stoddart. 2004. *Methods for the Economic Evaluation of Health Care Programmes*. Third. Oxford University Press.

Elliot, Steven, and David Kernick. 2011. "Why Do GPs with a Special Interest in Headache Investigate Headache Presentations with Neuroradiology and What Do They Find?" *The Journal of Headache and Pain* 12 (6): 625–28. https://doi.org/10.1007/s10194-011-0375-8.

EUnetHTA. 2015. "Methods_for_health_economic_evaluations." https://www.eunethta.eu/wp-content/uploads/2018/03/Methods_for_health_economic_evaluations.pdf.

European Society of Radiology. 2014. "Renewal of Radiological Equipment." *Insights into Imaging* 5 (5): 543–46. https://doi.org/10.1007/s13244-014-0345-1.

EuroQol Research Foundation. 2019. "EQ-5D-5L User Guide: Basic Information on How to Use the EQ-5D-5L Instrument." https://euroqol.org/wp-content/uploads/2019/09/EQ-5D-5L-English-User-Guide_version-3.0-Sept-2019-secured.pdf.

Fenwick, Elisabeth, Bernie J. O'Brien, and Andrew Briggs. 2004. "Cost-Effectiveness Acceptability Curves--Facts, Fallacies and Frequently Asked Questions." *Health Economics* 13 (5): 405–15. https://doi.org/10.1002/hec.903.

Fischer, Florian, Kerstin Lange, Kristina Klose, Wolfgang Greiner, and Alexander Kraemer. 2016. "Barriers and Strategies in Guideline Implementation—A Scoping Review." *Healthcare* 4 (3). https://doi.org/10.3390/healthcare4030036.

Fitzpatrick, R, and A Hopkins. 1981. "Referrals to Neurologists for Headaches Not Due to Structural Disease." *Journal of Neurology, Neurosurgery, and Psychiatry* 44 (12): 1061–67.

Franklin, Matthew, and Joanna Thorn. 2019. "Self-Reported and Routinely Collected Electronic Healthcare Resource-Use Data for Trial-Based Economic Evaluations: The Current State of Play in England and Considerations for the Future." *BMC Medical Research Methodology* 19 (1): 8. https://doi.org/10.1186/s12874-018-0649-9.

Freedman, L. S. 1987. "Evaluating and Comparing Imaging Techniques: A Review and Classification of Study Designs." *The British Journal of Radiology* 60 (719): 1071–81. https://doi.org/10.1259/0007-1285-60-719-1071.

Freemantle, Nick, Louise Marston, Kate Walters, John Wood, Matthew R. Reynolds, and Irene Petersen. 2013. "Making Inferences on Treatment Effects from Real World Data: Propensity Scores, Confounding by Indication, and Other Perils for the Unwary in Observational Research." *BMJ (Clinical Research Ed.)* 347 (November): f6409. https://doi.org/10.1136/bmj.f6409.

Frueh, Felix W., and Bruce Quinn. 2014. "Molecular Diagnostics Clinical Utility Strategy: A Six-Part Framework." *Expert Review of Molecular Diagnostics* 14 (7): 777–86. https://doi.org/10.1586/14737159.2014.933075.

Fryback, D. G., and J. R. Thornbury. 1991. "The Efficacy of Diagnostic Imaging." *Medical Decision Making: An International Journal of the Society for Medical Decision Making* 11 (2): 88–94.

Ganeshalingam, R, K Eng, and Rs Page. 2013. “Magnetic Resonance Imaging in the Acute Management of Suspected Scaphoid Fractures: A Review of the Literature and Assessment of Treatment Algorithm.” *OA Musculoskeletal Medicine* 1 (2). https://doi.org/10.13172/2052-9287-1-2-868.

Garattini, L, A Curto, A Padula, and N Freemantle. 2016. “Real-World Evidence in Economic Evaluations: Really Realistic?” *Journal of the Royal Society of Medicine* 109 (11): 404–7. https://doi.org/10.1177/0141076816671258.

Garrison, Louis P., Peter J. Neumann, Pennifer Erickson, Deborah Marshall, and C. Daniel Mullins. 2007. “Using Real-World Data for Coverage and Payment Decisions: The ISPOR Real-World Data Task Force Report.” *Value in Health: The Journal of the International Society for Pharmacoeconomics and Outcomes Research* 10 (5): 326–35. https://doi.org/10.1111/j.1524-4733.2007.00186.x.

Gatsonis, Constantine. 2012. “Standards in the Design, Conduct and Evaluation of Diagnostic Testing For Use in Patient Centered Outcomes Research.” https://www.pcori.org/assets/Standards-in-the-Design-Conduct-and-Evaluation-of-Diagnostic-Testing-for-Use-in-Patient-Centered-Outcomes-Research.pdf.

Gazelle, G.Scott, Larry Kessler, David W. Lee, Thomas McGinn, Joseph Menzin, Peter J. Neumann, Derek van Amerongen, and Leigh Ann White. 2011. “A Framework for Assessing the Value of Diagnostic Imaging in the Era of Comparative Effectiveness Research.” *Radiology* 261 (3): 692–98. https://doi.org/10.1148/radiol.11110155.

Glasgow, Russell E., Samantha M. Harden, Bridget Gaglio, Borsika Rabin, Matthew Lee Smith, Gwenndolyn C. Porter, Marcia G. Ory, and Paul A. Estabrooks. 2019. “RE-AIM Planning and Evaluation Framework: Adapting to New Science and Practice With a 20-Year Review.” *Frontiers in Public Health* 7. https://doi.org/10.3389/fpubh.2019.00064.

Glick, Henry, Jalpa Doshi, Seema Sonnad, and Daniel Polsky. 2007. *Economic Evaluation in Clinical Trials*. First Edition. Oxford University Press.

Gomes, Manuel, Robert W. Aldridge, Peter Wylie, James Bell, and Owen Epstein. 2013. “Cost-Effectiveness Analysis of 3-D Computerized Tomography Colonography Versus Optical Colonoscopy for Imaging Symptomatic Gastroenterology Patients.” *Applied Health Economics and Health Policy* 11 (2): 107–17. https://doi.org/10.1007/s40258-013-0019-z.

Gómez León, Nieves, Sofía Escalona, Beatriz Bandrés, Cristobal Belda, Daniel Callejo, and Juan Antonio Blasco. 2014. “(18)F-Fluorodeoxyglucose Positron Emission Tomography/Computed Tomography Accuracy in the Staging of Non-Small Cell Lung Cancer: Review and Cost-Effectiveness.” *Radiology Research and Practice* 2014: 135934. https://doi.org/10.1155/2014/135934.

Gonzalez, B, M Mahesh, KP Kim, M Bhargavan, R Lewis, F Mettler, and C Land. 2009. “Projected Cancer Risks from Computed Tomographic Scans Performed in the United States in 2007. - PubMed - NCBI.” https://www.ncbi.nlm.nih.gov/pubmed/20008689.

Good, Catriona. 2019. “GUIDELINES FOR NEUROIMAGING IN HEADACHE.” British Society of Neuroradiologists.

https://bsnr.org.uk/_userfiles/pages/files/standards_and_guidelines/bsnr_guidelines_for_imaging_in_headache_april_2019_final.pdf.

Gooding A., Coates M., and Rothwell A. 2004. “Cost Analysis of Traditional Follow-up Protocol versus MRI for Radiographically Occult Scaphoid Fractures: A Pilot Study for the Accident Compensation Corporation.” New Zealand Medical Journal. 2004. http://www.nzma.org.nz/journal/117-1201/1049/.

Gray, Alastair, Philip Clarke, Jane Wolstenholme, and Sarah Wordsworth. 2011. *Applied Methods of Cost-Effectiveness Analysis in Health Care*. First edition. Oxford University Press.

Grol, Richard, and Jeremy Grimshaw. 2003. “From Best Evidence to Best Practice: Effective Implementation of Change in Patients’ Care.” *Lancet (London, England)* 362 (9391): 1225–30. https://doi.org/10.1016/S0140-6736(03)14546-1.

Gupta, Sandeep K. 2011. “Intention-to-Treat Concept: A Review.” *Perspectives in Clinical Research* 2 (3): 109–12. https://doi.org/10.4103/2229-3485.83221.

Hackney, Lauren A, and Seth D Dodds. 2011. “Assessment of Scaphoid Fracture Healing.” *Current Reviews in Musculoskeletal Medicine* 4 (1): 16–22.

Halligan, Steve, Douglas G. Altman, Stuart A. Taylor, Susan Mallett, Jonathan J. Deeks, Clive I. Bartram, and Wendy Atkin. 2005. “CT Colonography in the Detection of Colorectal Polyps and Cancer: Systematic Review, Meta-Analysis, and Proposed Minimum Data Set for Study Level Reporting1.” *Radiology* 237 (3): 893–904. https://doi.org/10.1148/radiol.2373050176.

Halligan, Steve, Edward Dadswell, Kate Wooldrage, Jane Wardle, Christian von Wagner, Richard Lilford, Guiqing L. Yao, Shihua Zhu, and Wendy Atkin. 2015. “Computed Tomographic Colonography Compared with Colonoscopy or Barium Enema for Diagnosis of Colorectal Cancer in Older Symptomatic Patients: Two Multicentre Randomised Trials with Economic Evaluation (the SIGGAR Trials).” *Health Technology Assessment (Winchester, England)* 19 (54): 1–134. https://doi.org/10.3310/hta19540.

Halligan, Steve, Richard J. Lilford, Jane Wardle, Dion Morton, Pauline Rogers, Katherine Wooldrage, Rob Edwards, Reshma Kanani, Urvi Shah, and Wendy Atkin. 2007. “Design of a Multicentre Randomized Trial to Evaluate CT Colonography versus Colonoscopy or Barium Enema for Diagnosis of Colonic Cancer in Older Symptomatic Patients: The SIGGAR Study.” *Trials* 8 (1): 32. https://doi.org/10.1186/1745-6215-8-32.

Hanna, NJ, M Black, JW Sander, WH Smithson, R Appleton, S Brown, and DR Fish. 2002. “National Sentinel Clinical Audit of Epilepsy-Related Death.”

Hansen, Dane C., Sharat K. Kusuma, Ryan M. Palmer, and Kira B. Harris. 2014. “Robotic Guidance Does Not Improve Component Position or Short-Term Outcome in Medial Unicompartmental Knee Arthroplasty.” *The Journal of Arthroplasty* 29 (9): 1784–89.

Harvey, Gill, and Alison Kitson. 2015. “PARIHS Revisited: From Heuristic to Integrated Framework for the Successful Implementation of Knowledge into Practice.” *Implementation Science* 11 (1): 33. https://doi.org/10.1186/s13012-016-0398-2.

Hazard, Elisabeth, Julie Munakata, Marcelo E. Bigal, Marcia F. T. Rupnow, and Richard B. Lipton. 2009. “The Burden of Migraine in the United States: Current and Emerging Perspectives on Disease Management and Economic Analysis.” *Value in Health: The Journal of the International Society for Pharmacoeconomics and Outcomes Research* 12 (1): 55–64. https://doi.org/10.1111/j.1524-4733.2008.00404.x.

Higgins, JPT, and S Green. 2011. “Cochrane Handbook for Systematic Reviews of Interventions.” 2011. https://training.cochrane.org/handbook/current.

Hlatky, Mark A., Bernard De Bruyne, Gianluca Pontone, Manesh R. Patel, Bjarne L. Norgaard, Robert A. Byrne, Nick Curzen, et al. 2015. “Quality-of-Life and Economic Outcomes of Assessing Fractional Flow Reserve With Computed Tomography Angiography: PLATFORM.” *Journal of the American College of Cardiology* 66 (21): 2315–23. https://doi.org/10.1016/j.jacc.2015.09.051.

Hollingworth, William. 2005. “Radiology Cost and Outcomes Studies: Standard Practice and Emerging Methods.” *American Journal of Roentgenology* 185 (4): 833–39. https://doi.org/10.2214/AJR.04.1780.

Houn, Florence, Roselie A. Bright, Harry F. Bushar, Barbara Y. Croft, Charles A. Finder, John K. Gohagan, Robert J. Jennings, et al. 2000. “Study Design in the Evaluation of Breast Cancer Imaging Technologies.” *Academic Radiology* 7 (9): 684–92. https://doi.org/10.1016/S1076-6332(00)80524-3.

Howard, L, S Wessely, M Leese, L Page, P McCrone, K Husain, J Tong, and A Dowson. 2005. “Are Investigations Anxiolytic or Anxiogenic? A Randomised Controlled Trial of Neuroimaging to Provide Reassurance in Chronic Daily Headache.” *Journal of Neurology, Neurosurgery, and Psychiatry* 76 (11): 1558–64. https://doi.org/10.1136/jnnp.2004.057851.

Hu, X. Henry, Leona E. Markson, Richard B. Lipton, Walter F. Stewart, and Marc L. Berger. 1999. “Burden of Migraine in the United States: Disability and Economic Costs.” *Archives of Internal Medicine* 159 (8): 813–18. https://doi.org/10.1001/archinte.159.8.813.

Husereau, Don, Michael Drummond, Stavros Petrou, Chris Carswell, David Moher, Dan Greenberg, Federico Augustovski, Andrew H. Briggs, Josephine Mauskopf, and Elizabeth Loder. 2013. “Consolidated Health Economic Evaluation Reporting Standards (CHEERS) Statement.” *BMJ* 346 (March): f1049. https://doi.org/10.1136/bmj.f1049.

Jenkins, Paul J, Kate Slade, James S Huntley, and C Michael Robinson. 2008. “A Comparative Analysis of the Accuracy, Diagnostic Uncertainty and Cost of Imaging Modalities in Suspected Scaphoid Fractures.” *Injury* 39 (7): 768–74.

Jiang, Hongyu, and Xiao-Hua Zhou. 2004. “Bootstrap Confidence Intervals for Medical Costs with Censored Observations.” *Statistics in Medicine* 23 (21): 3365–76. https://doi.org/10.1002/sim.1556.

Jones, Andrew M. 2011. *Models for Health Care*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195398649.013.0024.

Kaewlai, Rathachai, Laura L Avery, Ashwin V Asrani, Hani H Abujudeh, Richard Sacknoff, and Robert A Novelline. 2008. “Multidetector CT of Carpal Injuries: Anatomy, Fractures, and Fracture-Dislocations.” *Radiographics: A Review Publication of the Radiological Society of North America, Inc* 28 (6): 1771–84. https://doi.org/10.1148/rg.286085511.

Karl, John W., Eric Swart, and Robert J. Strauch. 2015. “Diagnosis of Occult Scaphoid Fractures.” *J Bone Joint Surg Am* 97 (22): 1860–68. https://doi.org/10.2106/JBJS.O.00099.

Katkade, Vaibhav B, Kafi N Sanders, and Kelly H Zou. 2018. “Real World Data: An Opportunity to Supplement Existing Evidence for the Use of Long-Established Medicines in Health Care Decision Making.” *Journal of Multidisciplinary Healthcare* 11 (July): 295–304. https://doi.org/10.2147/JMDH.S160029.

Kelson, Tamika, Robert Davidson, and Tim Baker. 2016. “Early MRI versus Conventional Management in the Detection of Occult Scaphoid Fractures: What Does It Really Cost? A Rural Pilot Study.” *Journal of Medical Radiation Sciences* 63 (1): 9–16.

Kent, D. L., and E. B. Larson. 1992. “Disease, Level of Impact, and Quality of Research Methods. Three Dimensions of Clinical Efficacy Assessment Applied to Magnetic Resonance Imaging.” *Investigative Radiology* 27 (3): 245–54.

Kernick, David, and Stuart Williams. 2011. “Should GPs Have Direct Access to Neuroradiological Investigation When Adults Present with Headache?” *The British Journal of General Practice* 61 (587): 409–11. https://doi.org/10.3399/bjgp11X578124.

Kosinski, M., M. S. Bayliss, J. B. Bjorner, J. E. Ware, W. H. Garber, A. Batenhorst, R. Cady, C. G. H. Dahlöf, A. Dowson, and S. Tepper. 2003. “A Six-Item Short-Form Survey for Measuring Headache Impact: The HIT-6.” *Quality of Life Research: An International Journal of Quality of Life Aspects of Treatment, Care and Rehabilitation* 12 (8): 963–74. https://doi.org/10.1023/a:1026119331193.

Lang, Kathleen, Huan Huang, David W. Lee, Victoria Federico, and Joseph Menzin. 2013. “National Trends in Advanced Outpatient Diagnostic Imaging Utilization: An Analysis of the Medical Expenditure Panel Survey, 2000-2009.” *BMC Medical Imaging* 13 (1): 40. https://doi.org/10.1186/1471-2342-13-40.

Latinovic, R, M Gulliford, and L Ridsdale. 2006. “Headache and Migraine in Primary Care: Consultation, Prescription, and Referral Rates in a Large Population.” *Journal of Neurology, Neurosurgery, and Psychiatry* 77 (3): 385–87. https://doi.org/10.1136/jnnp.2005.073221.

Lee, David W., Richard Duszak, and Danny R. Hughes. 2013. “Comparative Analysis of Medicare Spending for Medical Imaging: Sustained Dramatic Slowdown Compared With Other Services.” *American Journal of Roentgenology* 201 (6): 1277–82. https://doi.org/10.2214/AJR.13.10999.

Lee, David W., Peter J. Neumann, and John A. Rizzo. 2010. “Understanding the Medical and Nonmedical Value of Diagnostic Testing.” *Value in Health: The Journal of the International Society for Pharmacoeconomics and Outcomes Research* 13 (2): 310–14. https://doi.org/10.1111/j.1524-4733.2009.00597.x.

Leonardi, M., and A. Raggi. 2013. “Burden of Migraine: International Perspectives.” *Neurological Sciences: Official Journal of the Italian Neurological Society and of the Italian Society of Clinical Neurophysiology* 34 Suppl 1 (May): S117-118. https://doi.org/10.1007/s10072-013-1387-8.

Lijmer, Jeroen G., Mariska Leeflang, and Patrick M. M. Bossuyt. 2009. “Proposals for a Phased Evaluation of Medical Tests.” *Medical Decision Making: An International Journal of the Society for Medical Decision Making* 29 (5): E13-21. https://doi.org/10.1177/0272989X09336144.

Linde, M., A. Gustavsson, L. J. Stovner, T. J. Steiner, J. Barré, Z. Katsarava, J. M. Lainez, et al. 2012. "The Cost of Headache Disorders in Europe: The Eurolight Project." *European Journal of Neurology* 19 (5): 703–11. https://doi.org/10.1111/j.1468-1331.2011.03612.x.

Loop, J. W., and L. E. Lusted. 1978. "American College of Radiology Diagnostic Efficacy Studies." *AJR. American Journal of Roentgenology* 131 (1): 173–79. https://doi.org/10.2214/ajr.131.1.173.

Low, G, and N Raby. 2005. "Can Follow-up Radiography for Acute Scaphoid Fracture Still Be Considered a Valid Investigation?" *Clinical Radiology* 60 (10): 1106–10. https://doi.org/10.1016/j.crad.2005.07.001.

Mackenzie, R., and A. K. Dixon. 1995. "Measuring the Effects of Imaging: An Evaluative Framework." *Clinical Radiology* 50 (8): 513–18. https://doi.org/10.1016/S0009-9260(05)83184-8.

MacPherson, Hugh. 2004. "Pragmatic Clinical Trials." *Complementary Therapies in Medicine* 12 (2–3): 136–40. https://doi.org/10.1016/j.ctim.2004.07.043.

Mallee, Wouter, Job N Doornberg, David Ring, C Niek van Dijk, Mario Maas, and J Carel Goslings. 2011. "Comparison of CT and MRI for Diagnosis of Suspected Scaphoid Fractures." *The Journal of Bone and Joint Surgery. American Volume* 93 (1): 20–28. https://doi.org/10.2106/JBJS.I.01523.

Manca, A., and P. C. Austin. 2008. "Using Propensity Score Methods to Analyse Individual Patient-Level Cost-Effectiveness Data from Observational Studies." 08/20. Health, Econometrics and Data Group (HEDG) Working Papers. HEDG, c/o Department of Economics, University of York. https://ideas.repec.org/p/yor/hectdg/08-20.html.

Manca, Andrea, Neil Hawkins, and Mark J. Sculpher. 2005. "Estimating Mean QALYs in Trial-Based Cost-Effectiveness Analysis: The Importance of Controlling for Baseline Utility." *Health Economics* 14 (5): 487–96. https://doi.org/10.1002/hec.944.

Marshall, Deborah A., and Margaret Hux. 2009. "Design and Analysis Issues for Economic Analysis alongside Clinical Trials." *Medical Care* 47 (7 Suppl 1): S14-20. https://doi.org/10.1097/MLR.0b013e3181a31971.

McCabe, Christopher, Karl Claxton, and Anthony J. Culyer. 2008. "The NICE Cost-Effectiveness Threshold." *PharmacoEconomics* 26 (9): 733–44. https://doi.org/10.2165/00019053-200826090-00004.

McCrone, Paul, Paul T. Seed, Andrew J. Dowson, Lucy V. Clark, Laura H. Goldstein, Myfanwy Morgan, and Leone Ridsdale. 2011. "Service Use and Costs for People with Headache: A UK Primary Care Study." *The Journal of Headache and Pain* 12 (6): 617–23. https://doi.org/10.1007/s10194-011-0362-0.

Meinecke, Anna-Katharina, Paco Welsing, George Kafatos, Des Burke, Sven Trelle, Maria Kubin, Gaelle Nachbaur, Matthias Egger, Mira Zuidgeest, and work package 3 of the GetReal consortium. 2017. "Series: Pragmatic Trials and Real World Evidence: Paper 8. Data Collection and Management." *Journal of Clinical Epidemiology* 91 (November): 13–22. https://doi.org/10.1016/j.jclinepi.2017.07.003.

Memarsadeghi, Mazda, Martin J Breitenseher, Cornelia Schaefer-Prokop, Michael Weber, Silke Aldrian, Christian Gäbler, and Mathias Prokop. 2006. “Occult Scaphoid Fractures: Comparison of Multidetector CT and MR Imaging--Initial Experience.” *Radiology* 240 (1): 169–76. https://doi.org/10.1148/radiol.2401050412.

Menardo, G. 2004. “Sensitivity of Diagnostic Examinations for Colorectal Polyps. - PubMed - NCBI.” *Tech Coloproctol*. https://www.ncbi.nlm.nih.gov/pubmed/15666105.

Moran, John L., Patricia J. Solomon, Aaron R. Peisach, and Jeffrey Martin. 2007. “New Models for Old Questions: Generalized Linear Models for Cost Prediction.” *Journal of Evaluation in Clinical Practice* 13 (3): 381–89. https://doi.org/10.1111/j.1365-2753.2006.00711.x.

Moreno Ramos, M. D., M. Martínez Hervás, P. Sanz Rupp, and J. Ramos Medrano. 2013. “[Analysis of the management of occult fractures of the scaphoid through early magnetic resonance imaging].” *Radiología* 55 (3): 247–52. https://doi.org/10.1016/j.rx.2011.06.012.

Morgan, Myfanwy, Linda Jenkins, and Leone Ridsdale. 2007. “Patient Pressure for Referral for Headache: A Qualitative Study of GPs’ Referral Behaviour.” *The British Journal of General Practice: The Journal of the Royal College of General Practitioners* 57 (534): 29–35.

National Institute for Health Research. 2009. “Special Report: Critical Appraisal of CT Colonography Cost-Effectiveness Analyses. - PubMed - NCBI.” *Technol Eval Cent Assess Program*. https://www.ncbi.nlm.nih.gov/pubmed/19824219.

Nguyen, Q, S Chaudhry, R Sloan, I Bhoora, and C Willard. 2008. “The Clinical Scaphoid Fracture: Early Computed Tomography as a Practical Approach.” *Annals of The Royal College of Surgeons of England* 90 (6): 488–91. https://doi.org/10.1308/003588408X300948.

NHS digital. 2018. “Prescription Cost Analysis - England, 2018 [PAS].” NHS Digital. 2018. https://digital.nhs.uk/data-and-information/publications/statistical/prescription-cost-analysis/2018.

NHS England. 2011. “Cancer Survival Rates ‘Threatened by Rising Cost.’” Nhs.Uk. December 12, 2011. https://www.nhs.uk/news/cancer/cancer-survival-rates-threatened-by-rising-cost/.

———. 2019. “NHS England » Long Term Plan Implementation.” 2019. https://www.england.nhs.uk/cancer/strategy/.

NHS Improvement. 2017. “Reference Costs | NHS Improvement.” 2017. https://improvement.nhs.uk/resources/reference-costs/#rc1718.

———. 2018. “Health Trust Specialist Services Reference Costs 2016-17.” 2018. https://data.gov.uk/dataset/03209d9b-fe35-4f85-acc9-590c26c0b4d3/health-trust-specialist-services-reference-costs-2016-17.

NHS Improvement Diagnostics. 2012. “Supporting Direct Access to Diagnostic Imaging for Cancer.”

NICE. 2011a. “CG131 Colorectal Cancer: NICE Guideline.” Guidance/Clinical Guidelines. 2011. http://publications.nice.org.uk/colorectal-cancer-cg131.

———. 2011b. “Diagnostic Assessment Programme Manual.”

———. 2011c. “Diagnostics Assessment Programme Manual.”

———. 2011d. “Medical Technologies Evaluation Programme.”

———. 2012a. “7 Assessing Cost Effectiveness | The Guidelines Manual | Guidance | NICE.” 2012. https://www.nice.org.uk/process/pmg6/chapter/assessing-cost-effectiveness.

———. 2012b. “Methods for the Development of NICE Public Health Guidance (Third Edition) | Guidance | NICE.” 2012. https://www.nice.org.uk/process/pmg4/chapter/introduction.

———. 2013. “Guide to the Methods of Technology Appraisal 2013.” 2013. https://www.nice.org.uk/process/pmg9/chapter/foreword.

———. 2014. “Overview | Colorectal Cancer: Diagnosis and Management | Guidance | NICE.” https://www.nice.org.uk/guidance/cg131.

———. 2015. “1 Recommendations Organised by Site of Cancer | Suspected Cancer: Recognition and Referral | Guidance | NICE.” 2015. https://www.nice.org.uk/guidance/ng12/chapter/1-Recommendations-organised-by-site-of-cancer#lower-gastrointestinal-tract-cancers.

———. 2016. “Quality Statement 1: Direct Access to Diagnostic Tests | Suspected Cancer | Quality Standards | NICE.” 2016. https://www.nice.org.uk/guidance/qs124/chapter/Quality-statement-1-Direct-access-to-diagnostic-tests.

———. 2018a. “Clinical Guideline; Colorectal Cancer: The Diagnosis and Management of Colorectal Cancer.”

———. 2018b. “Management of Headaches.” 2018. https://pathways.nice.org.uk/pathways/headaches.

Nilsen, Per. 2015. “Making Sense of Implementation Theories, Models and Frameworks.” *Implementation Science* 10 (1): 53. https://doi.org/10.1186/s13012-015-0242-0.

OECD. 2018. “OECD Statistics.” 2018. https://stats.oecd.org/Index.aspx?DataSetCode=HEALTH_PROC.

Office for National Statistics. 2019a. “Average Weekly Earnings in Great Britain - Office for National Statistics.” 2019. https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/bulletins/averageweeklyearningsingreatbritain/july2019.

———. 2019b. “Employee Earnings in the UK - Office for National Statistics.” 2019. https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/earningsandworkinghours/bulletins/annualsurveyofhoursandearnings/2018.

O’Sullivan, Amy K., David Thompson, and Michael F. Drummond. 2005. “Collection of Health-Economic Data Alongside Clinical Trials: Is There a Future for Piggyback Evaluations?” *Value in Health* 8 (1): 67–79. https://doi.org/10.1111/j.1524-4733.2005.03065.x.

Osumili, Beatrice, Paul McCrone, Sian Cousins, and Leone Ridsdale. 2018. “The Economic Cost of Patients With Migraine Headache Referred to Specialist Clinics.” *Headache: The Journal of Head and Face Pain* 58 (2): 287–94. https://doi.org/10.1111/head.13210.

Otero, Hansel J., Frank J. Rybicki, Dan Greenberg, and Peter J. Neumann. 2008. “Twenty Years of Cost-Effectiveness Analysis in Medical Imaging: Are We Improving?” *Radiology* 249 (3): 917–25. https://doi.org/10.1148/radiol.2493080237.

Parody, Elizabeth, Salvador Pedraza, María M. García-Gil, Carlos Crespo, Joaquín Serena, and Antoni Dávalos. 2015. “Cost–Utility Analysis of Magnetic Resonance Imaging Management of Patients with Acute Ischemic Stroke in a Spanish Hospital.” *Neurology and Therapy* 4 (1): 25–37. https://doi.org/10.1007/s40120-015-0029-x.

Patel, Nirav K., Nigel Davies, Zulfiquar Mirza, and Martin Watson. 2013. “Cost and Clinical Effectiveness of MRI in Occult Scaphoid Fractures: A Randomised Controlled Trial.” *Emergency Medicine Journal* 30 (3): 202–7. https://doi.org/10.1136/emermed-2011-200676.

Patterson, V H, and T F Esmonde. 1993. “Comparison of the Handling of Neurological Outpatient Referrals by General Physicians and a Neurologist.” *Journal of Neurology, Neurosurgery, and Psychiatry* 56 (7): 830.

Peacock, J., S Kerry, and R Balise. 2017. *Presenting Medical Statistics from Proposal to Publication*. Second. Oxford University Press.

Pearl, W. S. 1999. “A Hierarchical Outcomes Approach to Test Assessment.” *Annals of Emergency Medicine* 33 (1): 77–84.

Pertile, Paolo, Albino Poli, Lorenzo Dominioni, Nicola Rotolo, Elisa Nardecchia, Massimo Castiglioni, Massimo Paolucci, William Mantovani, and Andrea Imperatori. 2015. “Is Chest X-Ray Screening for Lung Cancer in Smokers Cost-Effective? Evidence from a Population-Based Study in Italy.” *Cost Effectiveness and Resource Allocation: C/E* 13: 15. https://doi.org/10.1186/s12962-015-0041-0.

Petrou, Stavros. 2012. “Rationale and Methodology for Trial-Based Economic Evaluation.” *Clinical Investigation* 2: 1191–1200.

Petrou, Stavros, and Alastair Gray. 2011. “Economic Evaluation alongside Randomised Controlled Trials: Design, Conduct, Analysis, and Reporting.” *The BMJ* 342 (April). https://doi.org/10.1136/bmj.d1548.

Pharoah, Paul D. P., Bernadette Sewell, Deborah Fitzsimmons, Hayley S. Bennett, and Nora Pashayan. 2013. “Cost Effectiveness of the NHS Breast Screening Programme: Life Table Model.” *BMJ* 346 (May): f2618. https://doi.org/10.1136/bmj.f2618.

Pickhardt, Perry J., Cesare Hassan, Steve Halligan, and Riccardo Marmo. 2011. “Colorectal Cancer: CT Colonography and Colonoscopy for Detection--Systematic Review and Meta-Analysis.” *Radiology* 259 (2): 393–405. https://doi.org/10.1148/radiol.11101887.

Pillai, Anand, and Manav Jain. 2005. “Management of Clinical Fractures of the Scaphoid: Results of an Audit and Literature Review.” *European Journal of Emergency Medicine: Official Journal of the European Society for Emergency Medicine* 12 (2): 47–51.

Polgreen, Linnea A., and John M. Brooks. 2012. “Estimating Incremental Costs with Skew: A Cautionary Note.” *Applied Health Economics and Health Policy* 10 (5): 319–29. https://doi.org/10.2165/11632430-000000000-00000.

Pooler, B. Dustin, David H. Kim, and Perry J. Pickhardt. 2017. “Extracolonic Findings at Screening CT Colonography: Prevalence, Benefits, Challenges, and Opportunities.” *American Journal of Roentgenology* 209 (1): 94–102. https://doi.org/10.2214/AJR.17.17864.

Rabeneck, Linda, Lawrence F Paszat, Robert J Hilsden, Refik Saskin, Des Leddin, Eva Grunfeld, Elaine Wai, Meredith Goldwasser, Rinku Sutradhar, and Therese A Stukel. 2008. “Bleeding and Perforation after Outpatient Colonoscopy and Their Risk Factors in Usual Clinical Practice.” *Gastroenterology* 135 (6): 1899–1906, 1906.e1. https://doi.org/10.1053/j.gastro.2008.08.058.

Raby, N. 2001. “Magnetic Resonance Imaging of Suspected Scaphoid Fractures Using a Low Field Dedicated Extremity MR System.” *Clinical Radiology* 56 (4): 316–20. https://doi.org/10.1053/crad.2000.0657.

Ramsey, Scott D., Richard J. Willke, Henry Glick, Shelby D. Reed, Federico Augustovski, Bengt Jonsson, Andrew Briggs, and Sean D. Sullivan. 2015. “Cost-Effectiveness Analysis alongside Clinical Trials II-An ISPOR Good Research Practices Task Force Report.” *Value in Health: The Journal of the International Society for Pharmacoeconomics and Outcomes Research* 18 (2): 161–72. https://doi.org/10.1016/j.jval.2015.02.001.

Ramsey, Scott, Richard Willke, Andrew Briggs, Ruth Brown, Martin Buxton, Anita Chawla, John Cook, et al. 2005. “Good Research Practices for Cost-Effectiveness Analysis alongside Clinical Trials: The ISPOR RCT-CEA Task Force Report.” *Value in Health: The Journal of the International Society for Pharmacoeconomics and Outcomes Research* 8 (5): 521–33. https://doi.org/10.1111/j.1524-4733.2005.00045.x.

Ranganathan, Priya, C. S. Pramesh, and Rakesh Aggarwal. 2016. “Common Pitfalls in Statistical Analysis: Intention-to-Treat versus per-Protocol Analysis.” *Perspectives in Clinical Research* 7 (3): 144–46. https://doi.org/10.4103/2229-3485.184823.

Rasmussen, B. K., R. Jensen, M. Schroll, and J. Olesen. 1991. “Epidemiology of Headache in a General Population--a Prevalence Study.” *Journal of Clinical Epidemiology* 44 (11): 1147–57.

Ratnasingham, Kumaran, Tammy Lo, Karim Jamal, Lavanya Varatharajan, Yasmin Tabbakh, Husein Kaderbhai, and Nicholas J. West. 2017. “The Role of Colonoscopy and CT Colonography in Patients Presenting with Symptoms of Constipation.” *The British Journal of Radiology* 90 (1073): 20160147. https://doi.org/10.1259/bjr.20160147.

Ridsdale, Leone, Lucy V Clark, Andrew J Dowson, Laura H Goldstein, Linda Jenkins, Paul McCrone, Myfanwy Morgan, and Paul T Seed. 2007. “How Do Patients Referred to Neurologists for Headache Differ from Those Managed in Primary Care?” *The British Journal of General Practice* 57 (538): 388–95.

Ridsdale, Leone, Jane Doherty, Paul McCrone, Paul Seed, and Headache Innovation and Evaluation Group. 2008. “A New GP with Special Interest Headache Service: Observational Study.” *The British Journal of General Practice: The Journal of the Royal College of General Practitioners* 58 (552): 478–83. https://doi.org/10.3399/bjgp08X319440.

Ring, David, and Santiago Lozano-Calderón. 2008. “Imaging for Suspected Scaphoid Fracture.” *The Journal of Hand Surgery* 33 (6): 954–57. https://doi.org/10.1016/j.jhsa.2008.04.016.

Rua, Tiago, David Parkin, Vicky Goh, Paul McCrone, and Sam Gidwani. 2017. “The Economic Evidence for Advanced Imaging in the Diagnosis of Suspected Scaphoid Fractures: Systematic Review of Evidence.” *Journal of Hand Surgery (European Volume)*, November, 1753193417742553. https://doi.org/10.1177/1753193417742553.

Rua, Tiago, Sanjay Vijayanathan, David Parkin, Vicky Goh, Paul McCrone, and Sam Gidwani. 2018. “Rationale and Design of the SMaRT Trial: A Randomised, Prospective, Parallel, Non-Blinded, One-Centre Trial to Evaluate the Use of Magnetic Resonance Imaging in Acute Setting in Patients Presenting with Suspected Scaphoid Fracture.” *Clinical Trials (London, England)* 15 (2): 120–29. https://doi.org/10.1177/1740774517748320.

Rubio-Valera, Maria, Mariona Pons-Vigués, María Martínez-Andrés, Patricia Moreno-Peral, Anna Berenguera, and Ana Fernández. 2014. “Barriers and Facilitators for the Implementation of Primary Prevention and Health Promotion Activities in Primary Care: A Synthesis through Meta-Ethnography.” *PloS One* 9 (2): e89554. https://doi.org/10.1371/journal.pone.0089554.

Rycroft-Malone, Jo. 2004. “The PARIHS Framework—A Framework for Guiding the Implementation of Evidence-Based Practice:” *Journal of Nursing Care Quality* 19 (4): 297–304. https://doi.org/10.1097/00001786-200410000-00002.

Saxena, P, Ruth McDonald, S Gull, and N Hyder. 2003. “Diagnostic Scanning for Suspected Scaphoid Fractures: An Economic Evaluation Based on Cost-Minimisation Models.” *Injury* 34 (7): 503–11.

Schouw, Y. T. van der, A. L. Verbeek, and S. H. Ruijs. 1995. “Guidelines for the Assessment of New Diagnostic Tests.” *Investigative Radiology* 30 (6): 334–40.

Sculpher, Mark J., Karl Claxton, Mike Drummond, and Chris McCabe. 2006. “Whither Trial-Based Economic Evaluation for Health Care Decision Making?” *Health Economics* 15 (7): 677–87. https://doi.org/10.1002/hec.1093.

Sempere, A. P., J. Porta-Etessam, V. Medrano, I. Garcia-Morales, L. Concepción, A. Ramos, I. Florencio, F. Bermejo, and C. Botella. 2005. “Neuroimaging in the Evaluation of Patients with Non-Acute Headache.” *Cephalalgia: An International Journal of Headache* 25 (1): 30–35. https://doi.org/10.1111/j.1468-2982.2004.00798.x.

Silberstein, S D, and R B Lipton. 1993. “Epidemiology of Migraine.” *Neuroepidemiology* 12 (3): 179–94.

Silva, Debra. 2015. “What’s Getting in the Way? Barriers to Improvement in the NHS.” The Health Foundation.

Silverman, Stuart L. 2009. “From Randomized Controlled Trials to Observational Studies.” *The American Journal of Medicine* 122 (2): 114–20. https://doi.org/10.1016/j.amjmed.2008.09.030.

Silverstein, M. D., and B. J. Boland. 1994. “Conceptual Framework for Evaluating Laboratory Tests: Case-Finding in Ambulatory Patients.” *Clinical Chemistry* 40 (8): 1621–27.

Simpson, Graeme C., Kirsten Forbes, Evelyn Teasdale, Alok Tyagi, and Celestine Santosh. 2010. "Impact of GP Direct-Access Computerised Tomography for the Investigation of Chronic Daily Headache." *Br J Gen Pract* 60 (581): 897–901. https://doi.org/10.3399/bjgp10X544069.

Sistrom, Christopher L. 2009. "The Appropriateness of Imaging: A Comprehensive Conceptual Framework." *Radiology* 251 (3): 637–49. https://doi.org/10.1148/radiol.2513080636.

Sommerbakk, Ragni, Dagny Faksvåg Haugen, Aksel Tjora, Stein Kaasa, and Marianne Jensen Hjermstad. 2016. "Barriers to and Facilitators for Implementing Quality Improvements in Palliative Care – Results from a Qualitative Interview Study in Norway." *BMC Palliative Care* 15 (July). https://doi.org/10.1186/s12904-016-0132-5.

Sorenson, Corinna, Michael Drummond, and Beena Bhuiyan Khan. 2013. "Medical Technology as a Key Driver of Rising Health Expenditure: Disentangling the Relationship." *ClinicoEconomics and Outcomes Research: CEOR* 5 (May): 223–34. https://doi.org/10.2147/CEOR.S39634.

Stewart, W. F., R. B. Lipton, A. J. Dowson, and J. Sawyer. 2001. "Development and Testing of the Migraine Disability Assessment (MIDAS) Questionnaire to Assess Headache-Related Disability." *Neurology* 56 (6 Suppl 1): S20-28. https://doi.org/10.1212/wnl.56.suppl_1.s20.

Stovner, Lj, K Hagen, R Jensen, Z Katsarava, Rb Lipton, Ai Scher, Tj Steiner, and J-A Zwart. 2007. "The Global Burden of Headache: A Documentation of Headache Prevalence and Disability Worldwide." *Cephalalgia: An International Journal of Headache* 27 (3): 193–210. https://doi.org/10.1111/j.1468-2982.2007.01288.x.

Sudsawad, Pimjai. 2007. "Center on Knowledge Translation for Disability and Rehabilitation Research (KTDRR) KT Library - Research Quality." August 1, 2007. https://ktdrr.org/ktlibrary/articles_pubs/ktmodels/.

Symvoulakis, Emmanouil K, Lucy V Clark, Andrew J Dowson, Roger Jones, and Leone Ridsdale. 2007. "Headache: A 'suitable Case' for Behavioural Treatment in Primary Care?" *The British Journal of General Practice* 57 (536): 231–37.

Tal, Joseph. 2011. *Strategy and Statistics in Clinical Trials: A Non-Statistician's Guide to Thinking, Designing and Executing*. Academic Press.

Taylor, Timothy, Nikos Evangelou, Hugh Porter, William Hamilton, and David Kernick. 2014. "Headache: Two Views on the Right Approach in General Practice." *British Journal of General Practice* 64 (626): 475–76. https://doi.org/10.3399/bjgp14X681541.

Thom, Howard, Nicholas E J West, Vikki Hughes, Matthew Dyer, Martin Buxton, Linda D Sharples, Christopher H Jackson, Andrew M Crean, and The CECaT study group. 2014. "Cost-Effectiveness of Initial Stress Cardiovascular MR, Stress SPECT or Stress Echocardiography as a Gate-Keeper Test, Compared with Upfront Invasive Coronary Angiography in the Investigation and Management of Patients with Stable Chest Pain: Mid-Term Outcomes from the CECaT Randomised Controlled Trial." *BMJ Open* 4 (2): e003419. https://doi.org/10.1136/bmjopen-2013-003419.

Thomas, Ralph, Alan Cook, Gavin Main, Tom Taylor, Elizabeth Galizia Caruana, and Robert Swingler. 2010. "Primary Care Access to Computed Tomography for Chronic Headache." *The British Journal of*

General Practice: The Journal of the Royal College of General Practitioners 60 (575): 426–30. https://doi.org/10.3399/bjgp10X502146.

Tiel-van Buul, M M, T H Broekhuizen, E J van Beek, and P M Bossuyt. 1995. "Choosing a Strategy for the Diagnostic Management of Suspected Scaphoid Fracture: A Cost-Effectiveness Analysis." *Journal of Nuclear Medicine : Official Publication, Society of Nuclear Medicine* 36 (1): 45–48.

Titler, Marita. 2018. "Translation Research in Practice: An Introduction." *The Online Journal of Issues of Nursing* 2.

Tsushima, Yoshito, and Keigo Endo. 2005. "MR Imaging in the Evaluation of Chronic or Recurrent Headache." *Radiology* 235 (2): 575–79. https://doi.org/10.1148/radiol.2352032121.

Underwood, Raphael, Rachael Kilner, and Leone Ridsdale. 2017. "Primary Care Management of Headaches and How Direct-Access MRI Fits: A Qualitative Study of UK General Practitioners' Views." *BMJ Open* 7 (11): e018169. https://doi.org/10.1136/bmjopen-2017-018169.

Vos, Theo, Abraham D. Flaxman, Mohsen Naghavi, Rafael Lozano, Catherine Michaud, Majid Ezzati, Kenji Shibuya, et al. 2012. "Years Lived with Disability (YLDs) for 1160 Sequelae of 289 Diseases and Injuries 1990–2010: A Systematic Analysis for the Global Burden of Disease Study 2010." *The Lancet* 380 (9859): 2163–96. https://doi.org/10.1016/S0140-6736(12)61729-2.

Wallace, John. 2013. "Lost in Translation: Transferring Knowledge from Research to Clinical Practice." *Advances in Psychiatric Treatment* 19: 250–58.

Wang, Rongfei, Ruozhuo Liu, Zhao Dong, Hui Su, Ran Ao, Yinglu Liu, Yan Wang, Lin Ma, and Shengyuan Yu. 2019. "Unnecessary Neuroimaging for Patients With Primary Headaches." *Headache* 59 (1): 63–68. https://doi.org/10.1111/head.13397.

Weintraub, William S., Thomas F. Lüscher, and Stuart Pocock. 2015. "The Perils of Surrogate Endpoints." *European Heart Journal* 36 (33): 2212–18. https://doi.org/10.1093/eurheartj/ehv164.

Welsing, Paco M., Katrien Oude Rengerink, Sue Collier, Laurent Eckert, Maarten van Smeden, Antonio Ciaglia, Gaelle Nachbaur, et al. 2017. "Series: Pragmatic Trials and Real World Evidence: Paper 6. Outcome Measures in the Real World." *Journal of Clinical Epidemiology* 90 (October): 99–107. https://doi.org/10.1016/j.jclinepi.2016.12.022.

Wensing, Michel, and Richard Grol. 2019. "Knowledge Translation in Health: How Implementation Science Could Contribute More." *BMC Medicine* 17 (1): 88. https://doi.org/10.1186/s12916-019-1322-9.

Whicher, Danielle M, Jennifer E Miller, Kelly M Dunham, and Steven Joffe. 2015. "Gatekeepers for Pragmatic Clinical Trials." *Clinical Trials: Journal of the Society for Clinical Trials* 12 (5): 442–48. https://doi.org/10.1177/1740774515597699.

White, Ian R., Patrick Royston, and Angela M. Wood. 2011. "Multiple Imputation Using Chained Equations: Issues and Guidance for Practice." *Statistics in Medicine* 30 (4): 377–99. https://doi.org/10.1002/sim.4067.

Wood, Richard. 2018. “Oxfordshire Headache Pathway for the Efficient Diagnostic and Management Support of Headache Disorders.” SharedLearningArticle. NICE. 2018. https://www.nice.org.uk/sharedlearning/oxfordshire-headache-pathway-for-the-efficient-diagnostic-and-management-support-of-headache-disorders.

Woolacott, Nerys, Mark Corbett, Julie Jones-Diette, and Robert Hodgson. 2017. “Methodological Challenges for the Evaluation of Clinical Effectiveness in the Context of Accelerated Regulatory Approval: An Overview.” *Journal of Clinical Epidemiology* 90 (October): 108–18. https://doi.org/10.1016/j.jclinepi.2017.07.002.

Xiong, T., M. Richardson, R. Woodroffe, S. Halligan, D. Morton, and R. J. Lilford. 2005. “Incidental Lesions Found on CT Colonography: Their Nature and Frequency.” *The British Journal of Radiology* 78 (925): 22–29. https://doi.org/10.1259/bjr/67998962.

Yang, Szu-Chun, Wu-Wei Lai, Chien-Chung Lin, Wu-Chou Su, Li-Jung Ku, Jing-Shiang Hwang, and Jung-Der Wang. 2017. “Cost-Effectiveness of Implementing Computed Tomography Screening for Lung Cancer in Taiwan.” *Lung Cancer (Amsterdam, Netherlands)* 108: 183–91. https://doi.org/10.1016/j.lungcan.2017.04.001.

Yin, Zhong-Gang, Jian-Bing Zhang, and Ke-Tong Gong. 2015. “Cost-Effectiveness of Diagnostic Strategies for Suspected Scaphoid Fractures.” *Journal of Orthopaedic Trauma*, March. https://doi.org/10.1097/BOT.0000000000000316.

Yin, Zhong-Gang, Jian-Bing Zhang, Shi-Lian Kan, and Xiao-Gang Wang. 2010. “Diagnosing Suspected Scaphoid Fractures: A Systematic Review and Meta-Analysis.” *Clinical Orthopaedics and Related Research* 468 (3): 723–34. https://doi.org/10.1007/s11999-009-1081-6.

Zhou, Alice, David M. Yousem, and Matthew D. Alvin. 2018. “Cost-Effectiveness Analysis in Radiology: A Systematic Review.” *Journal of the American College of Radiology* 15 (11): 1536–46. https://doi.org/10.1016/j.jacr.2018.06.018.

Zoellner, Jamie M., and Kathleen J. Porter. 2017. “Chapter 6 - Translational Research: Concepts and Methods in Dissemination and Implementation Research.” In *Nutrition in the Prevention and Treatment of Disease (Fourth Edition)*, edited by Ann M. Coulston, Carol J. Boushey, Mario G. Ferruzzi, and Linda M. Delahanty, 125–43. Academic Press. https://doi.org/10.1016/B978-0-12-802928-2.00006-0.

Zoellner, Jamie, Linda Van Horn, Philip M. Gleason, and Carol J. Boushey. 2015. “What Is Translational Research? Concepts and Applications in Nutrition and Dietetics.” *Journal of the Academy of Nutrition and Dietetics* 115 (7): 1057–71. https://doi.org/10.1016/j.jand.2015.03.010.